THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Exploring the Mechanisms of Fibrillar Protein Aggregation

Ryan J. Morris

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy
to the
University of Edinburgh
September 2013

Abstract

The aim of this thesis is to investigate and better understand the mechanisms of protein self-assembly. Specifically, I study three protein systems which form morphologically and structurally distinct fibrillar protein aggregates. The first of these studies is concerned with the self-assembly of amyloid fibrils formed from bovine insulin. Amyloid fibrils are associated with human diseases such as Alzheimers Disease and type-2 diabetes, and are also garnering interest in biomaterial applications. Fragmentation-dominated models for the self-assembly of amyloid fibrils have had important successes in explaining the kinetics of amyloid fibril formation but predict fibril length distributions that do not match experimental observations. Here I resolve this inconsistency using a combination of experimental kinetic measurements and computer simulations. I provide evidence for a structural transition demarcated by a critical fibril mass concentration, or CFC, above which fragmentation of the fibrils is suppressed. Our simulations predict the formation of distinct fibril length distributions above and below the CFC, which I confirm by electron microscopy. These results point to a new picture of amyloid fibril growth in which structural transitions that occur during self-assembly have strong effects on the final population of aggregate species with small, and potentially cytotoxic, oligomers dominating for long periods of time at protein concentrations below the CFC. I further show that the CFC can be modulated by environmental conditions, pointing to possible *in vivo* strategies for controlling cytotoxicity. I probe the structural nature of the transition by performing small angle neutron scattering.

Secondly, I study the formation of amyloid-like fibrils from the protein ovalbumin. I undertake kinetic experiments of self-assembly and find two key features emerge: the lack of a lag time and the existence of a slow growth regime in the long-time limit. I observe, using TEM, that these fibrils are worm-like in

nature and form closed-loops. I find the growth kinetics are intimately connected to this particular morphology. I present a simple kinetic model which captures the features of the kinetics found in experiments by incorporating end-to-end association of fibrils. I comment on the ramifications this type of amyloid fibril assembly may have on oligomeric toxicity.

Thirdly, the DNA-mimic protein ocr is highly charged (-56e at pH 8) and forms non-amyloid fibrillar assemblies at very high ammonium sulphate concentrations (3.2M). The fact that ocr forms translucent fibrillar gels at such high salt concentrations is extremely unique. Typically under such high salt conditions, non-specific amorphous aggregates are formed. In order to better understand the mechanism of why ocr forms specific fibril aggregates, I used variants of the wile-type protein in which extensive regions of surface have been removed or modified. The structural characteristics of gels formed from the variants were probed using microrheological techniques. I find that non-specific electrostatic charge screening plays an important role in ocr aggregation. However, I also locate a potentially important α-helical region which may play a part in establishing specific interactions so that ocr may form ordered fibrillar assemblies.

Lay Summary

Proteins are complex biological molecules which act as the primary work horse for the cell and can be found in every tissue and organ within the body. They perform many critical functions that maintain life such as replicating DNA, transporting small molecules around the cell, or providing structural rigidity to cells and tissues. Their ability to do their job properly is often intimately connected to their three-dimensional shape, often referred to as the 'native fold'. However, in many circumstances, the native fold can become disrupted and misfolded or partially unfolded. When 'sticky' regions on the protein become exposed, which otherwise would be hidden when the protein is in its native fold this, quite often proteins like to join together. This process is known as protein aggregation. There are many structures that aggregates can form, from amorphous blobs, to structured filaments or fibrils.

Scientists are interested in how and why protein aggregation occurs for many reasons. One extremely important purpose for studying protein aggregation is that such processes are involved in many human diseases. Well known examples of diseases associated with protein aggregation are Alzheimer's disease, Parkinson's disease, and Type-2 diabetes. Furthermore, protein-based pharmaceuticals can be very susceptible to aggregation. If protein-based drugs aggregate, they are rendered useless. It is therefore of great importance in the pharmaceutical industry to understand, and in turn, prevent protein aggregation during the production and storage of pharmaceuticals to ensure stable drugs with long shelf lives. However, protein aggregation is not always entirely unwanted. Many researchers are now becoming increasingly interested in using certain types of protein aggregates as the basis for new and exciting materials.

In this thesis I investigate the basic mechanisms which drive the assembly of proteins into fibrillar aggregates. I use three different types of proteins which

all form structurally different filamentous aggregates. First, insulin forms a particular type of aggregate known as an amyloid fibril. These aggregates are closely associated with the diseases mentioned above. The important result from this work is that I found conditions in which the potentially harmful aggregates which may cause these diseases can either drastically accumulate or be effectively minimized due to structural changes of the fibrillar aggregates. This insight may have ramifications for understanding the progression of these diseases within the body.

Second, I use another protein, ovalbumin, which forms flexible, worm-like, fibrillar aggregates. I find that these aggregates can form circular fibrils or loops. By using a simple mathematical model, I am able to show that interesting features of how they grow can be successfully explained by their ability to form these loops. The mechanism of end-joining could provide a different mechanism from the one mentioned above on how to prevent the accumulation of harmful aggregates associated with disease.

Finally, I study another protein, called ocr, that forms a highly unusual fibril aggregate. Typically, proteins in very salty solutions aggregate, but they almost always form amorphous, globular aggregates. Ocr is unique in that it forms fibrils in the presence of high concentrations of salt. I used genetically modified versions of the protein, and studied the structure of the subsequent aggregate gel that is formed, to understand how and why this protein surprisingly forms this type of aggregate under these conditions.

Declaration

Except where otherwise stated, the research undertaken in this thesis was the unaided work of the author. Where the work was done in collaboration with others, a significant contribution was made by the author.

R.J. Morris

September 2013

Acknowledgements

First and foremost I'd like to express my extreme gratitude to Prof. Cait MacPhee for being an exemplary supervisor. Without her continual encouragement, patience, understanding and guidance I would not have been able to complete this thesis.

I also must thank Dr. Rosalind Allen for spending many hours discussing portions of the work contained in this thesis. I all too often stuck my head into her office late in the day to discuss an issue or a new idea. She always had time, despite her busy schedule, to engage with me and our discussions always proved extremely helpful.

The neutron scattering experiments would not have been possible without the knowledge and expertise of Dr. Simon Titmuss. I am greatly appreciative for his proof reading and helpful suggestions concerning that chapter.

I need to thank Dr. David Dryden and his group, particularly Dr. Augustinos Stephanou, Dr. Gareth Roberts and Mr. Laurie Cooper, in the School of Chemistry for providing ocr and the mutants used in this thesis.

Many thanks go to the members of my thesis committee, Dr. Tuomas Knowles, Dr. Frank Gunn-Moore, and Dr. Davide Marenduzzo, for taking the time to not only read this work but engaging me in a very stimulating and edifying *viva voce*.

A large thank you must go to Mr. Kym Eden. He was the first person I met when I moved to Edinburgh. It was a great pleasure to find that we would be working closely together on these projects. Much of the work in this thesis has been achieved through long hours of discussion and collaboration between the two of us and I certainly would not have been able to produce this thesis without his hard work, knowledge, and enthusiasm.

I'd also like to acknowledge the students that have helped me with some of the experiments that are contained in this thesis. Firstly, Reuben Yarwood, who

worked on the kinetics of insulin amyloid fibril formation in the presence of NaCl. Second, Sean Tabor who worked on the ovalbumin kinetic experiments. Finally, Sam Covill who helped me with performing the Bradford Assay and ovalbumin temperature experiments.

I would like to say thank you to all my fellow PhD. students that I have spent time with here. You have made it an incredibly enjoyable and edifying experience. I'd also like to thank Jane Patterson as she has been an immense help when dealing with the seemingly copious amounts of paper work needing to be done when you are a non-native student. I would also like to acknowledge all those who attend the monthly amyloid meetings for the often lively discussion and debates, particularly Dr. Jason Kalapothakis and Dr. Perdita Barran.

I of course need to thank my mother, Gail, and father, Evan, for their unwavering support over the years. I also need to express my deepest gratitude to my grandparents John and Francesca, for their support and encouragement. My grandmother's continual question of when will I go back and do my PhD. helped spur me on to finally achieve this long held goal.

I'd be remiss if I did not mention Shannon and Damian Alexander for their generosity and friendship.

Above all else, I thank my wife Ivy for her incredible love and support. We have uprooted our lives, left the city we loved and knew so well, to forge a new and uncertain life here in a far away place. That you took that gamble with me I shall be forever thankful.

Ultimately, I dedicate this thesis to Skye, our first child who was born just prior to this thesis being completed. You inspire me to do good work and I hope I can make you proud.

Contents

List of Figures

Chapter 1

Introduction

1.1 Protein Folding

Proteins are ubiquitous biological macromolecules in living systems; they can be found in every cell and every component of the cell. All proteins, at their most basic level, are biopolymer chains comprised from a set of 20 amino acids. These subunits are covalently linked together by the characteristic peptide bond. From this starting point, a startling diversity in protein sizes, structures, and functions emerges. Proteins play a role in various biochemical mechanisms, from transmembrane ion transport, carbohydrate metabolism, controlling cell cycle and differentiation, to a myriad of other functions [3].

Such diversity and the specificity of their function is often directly linked to their three dimensional native structure. How a polypeptide chain self-assembles from an unstructured polymer to a biologically viable three dimensional protein is one of the fundamental questions of life. Anfinsen's pioneering work [11, 12] showed that proteins self-assemble almost spontaneously, without any external input. In a seminal experiment he showed that urea unfolded ribonuclease can be restored to its native conformation by returning the protein to physiological conditions. In Anfinsen's own words this reversibility shows "that the information...for the assumption of the native secondary and tertiary structures, is contained in the amino acid sequence itself"[11]. This is the crux of what has become known as 'Anfinsen's Dogma': under physiological conditions a protein will adopt a configuration of the amino acid sequence that is a minimum of its free energy landscape. While this may explain why proteins fold it does not

explain *how* that fold is achieved. An immediate problem arises: the probability of finding the native fold by randomly searching the configuration space of even a moderately sized polypeptide chain ($\sim$ 100 amino acids in length) results in folding times longer than the age of the universe [13, 14]. The folding time of most proteins is on the order of milliseconds (some even fold on the microsecond time scale [15]). Clearly, proteins can not randomly search the configuration space to find their native state; some more efficient mechanism must govern protein folding. One solution to this problem, argued by Levinthal, was that there are specific 'folding pathways' which are guided by local interactions between specific amino acid groups creating stable, native-like structures. These interactions act effectively as nucleation steps in the folding process [16].

With the advent of techniques which allow for greater molecular resolution (i.e. NMR) and computational power, the kinetics of folding has shifted to a 'new' view of protein folding. Protein folding does not seem to require specific, step-wise pathways to achieve the native fold; rather it is a stochastic process which relies on the inherent fluctuations in the conformation of the unfolded state. What has emerged is a picture of trial-and-error: stochastic fluctuations of residues in a polypeptide chain can make native-like contacts, funneling the chain down an energy landscape to its native state [17]. Indeed, if the energy landscape is of the right shape, only a small number of contacts need to be made in order for the chain to adopt the correct native state [18].

1.2 Protein Aggregation

There are inevitably situations, both *in vivo* and *in vitro*, in which the native and biologically functional conformational state of a protein is not achieved or becomes perturbed. When a protein becomes misfolded or partially unfolded, multiple protein units can self-associate into larger, often insoluble and irreversible, complexes via non-native interactions. This process is known as protein aggregation. Often in protein aggregation, the non-native contacts established between protein units are driven by inter-protein hydrophobic interactions between exposed hydrophobic regions on misfolded proteins [19]. These non-native states can arise due to thermal or oxidative stresses [20, 21], mutations of amino acid sequences which may lead to accumulation of kinetically

trapped, partially unfolded intermediates or to misfolded states intrinsically prone to aggregation[20], the presence of hydrophobic surfaces and alterations of solution conditions such as pH (the optimal pH for stability of the native structure is 7.2-7.4) and the presence of salts and other co-solutes [22].

1.2.1 Multiple pathways of protein aggregation leads to a diversity of aggregate morphologies.

There are a variety of pathways in which proteins may aggregate with varying morphological and structural end products. These pathways may be dependent upon the type of protein, the initial conformational state of the protein, or the environmental conditions. Broadly speaking, protein aggregates can broken down into two classes: ordered and disordered.

Ordered protein aggregates are typified by the presence of long-range structural organization within the aggregate complex. The specificity of the interactions which give rise to this long-range order can induce a directionality in the assembly of the aggregate resulting in filamentous or fibrillar structures. The most well known example of this class of aggregate are the amyloid fibrils. Generically, amyloid fibrils are filamentous aggregates with large proportions of β-sheet comprising the structure of the fibrils. This preponderance of β-sheet within the fibril arises regardless of the initial native conformation of the constituent protein. Amyloid fibrils, their structure and formation will be discussed in further detail below. There are many ordered protein fibril aggregate assemblies which do not bear the characteristics of amyloid fibrils–for instance the fibrillar polymerization of haemoglobin in erythrocytes affected by sickle cell disease [23].

Disordered, amorphous aggregates are characterized by the non-specific agglomeration of protein into complexes without any discernible repetitive or long-range structural characteristics. An archetypal example of this is the heat denaturation of hen egg white. Another example are protein precipitates formed during an ammonium sulphate 'salting out' step in a protein purification procedure. In that case, the addition of salt ions to a protein solution causes a dehydration of the solvation shell which surrounds the protein. This loss of the interaction of water molecules from the charged protein surface allows for protein-protein hydrophobic interactions to become important. These non-specific hydrophobic interactions induce proteins to aggregate into clusters that

eventually become larger enough to be seen by eye and may precipitate out of solution in amorphous flocs. It should be noted that certain salt anions and cations are more efficient at removing the hydration from the protein surface and thus are better at precipitating proteins from solution. This hierarchy of cations and anions is known as the Hofmeister series.

Straddling these two extremes are the spherical or particulate protein aggregates which appear 'regular' in structure at large length scales (i.e. larger than protein molecules) but are disordered and fractal-like at shorter length scales. Many proteins have been observed to form these types of structures such as β-lactoglobulin, bovine serum albumin, hen egg white lysozyme, and casein [24, 25, 22]. Unlike amyloid fibrils, they lack a common structural motif and often contain structure similar to the native conformation of the constituent protein. It has been observed that for many proteins under partially denaturing conditions, these particulate protein aggregates often form near their isoelectric point (pI) or under conditions of large salt-screening [22]. Intriguingly, Krebs *et al.* found that the same proteins will form amyloid fibrils when the pH of the solution was changed so that the protein possessed a high positive net charge. The picture that emerges is that, under partially denaturing conditions, solution conditions, and in turn the surface charge, will determine to a large extent the aggregate morphology. The lack of electrostatic repulsion at the pI or high ionic strength allows for rapid and non-specific aggregation; this lack of specificity results in an absence of directionality in the aggregate formation resulting in spherical aggregate particles. In contrast, highly charged partially unfolded aggregates will associate slowly due to electostatic repulsion. This allows time for specific structural re-organization into β-sheets. This structural specificity imparts the uni-directional nature of the aggregation process which results in linear aggregates. It is clear that the conditions in which aggregates form is critical in determining the pathway of formation and subsequent morphology.

If protein concentrations are sufficiently high, a protein aggregate gel may be formed whose structure is comprised of the various aggregates discussed above. The underlying aggregate morphology of a protein gel will, to a large extent, determine its rheological and optical properties. It was observed that transparent protein gels were composed of a network of fine-stranded aggregates with a filament diameter on the order of the native protein diameter [26, 27, 28]. Turbid

and opaque protein gels were found to form from colloidal protein aggregates or coarsened fibrillar aggregates with diameters $> 0.5\mu$m [26, 27, 28]. In light of the previous discussion, electrostatic charge is crucial in determining the properties of these gels. For pH values greater than or less than the pI, fine-stranded gels are formed, but often with distinct rheological properties. When the pH $\sim$ pI, turbid, particulate gels are formed. Such control over structural and optical properties are important in the food sciences where mouth-feel, texture, and appearance will be markedly different for the same protein gel made under different conditions (e.g. whey protein isolate gels) [29, 30]. It is clear that the structure of a protein gel is inherently linked with the underlying protein aggregate morphology.

1.2.2 Aggregation *in vivo*

Evolution has equipped cells with biological 'machinery' to guide the protein assembly process and to sequester and destroy proteins that are not correctly folded. One important class of molecular objects that aid in this process are the so-called chaperones proteins. Chaperones can play various roles, from interacting with newly synthesized polypeptide chains to aiding in the final steps of protein folding. Chaperones do not act as catalysts, but instead ensure efficient and faithful assembly of polypeptide chains into properly folded proteins. Well studied examples of these are the family of heat shock proteins (HSP), in particular the GroEL chaperone. GroEL can sequester partially unfolded regions of a polypeptide chain from the surrounding environment to ensure the adoption of proper native structure [31].

There exists other 'quality-control' mechanisms within the cell, particularly in the endoplasmic reticulum (ER) where many proteins are translocated to fold properly, before secretion. In addition to the presence of a wide range of chaperones, folding catalysts are present to ensure efficient and expedient folding. Such quality-control is achieved via specific glycosylation and de-glycosylation of proteins, allowing the cell to distinguish between folded and misfolded proteins [32]. If a protein is recognized through this process as being misfolded it is ubiquitinated and trafficked to the proteasome to be safely degraded.

Since proteins are the molecules that drive most processes in life, misfolded proteins can obviously result in biological malfunction and therefore disease. One way in which misfolded proteins result in disease is when there are genetic

errors which encode for certain proteins. For instance, many misfolded protein pathologies are characterized by an absence of a particular protein. This is due to the efficient and effective mechanisms of protein folding quality control: as mutant protein generated from genetic defects emerges from the ribosome and into the ER, it will often fold improperly. The quality control mechanism of the cell takes over and it is safely eliminated. Many protein misfolding diseases originate in this way: cystic fibrosis (CFTR protein), Marfan syndrome (fibrillin), Fabry disease (α-galactosidase) and Gauchers disease (misfolded β-glucocerebrosidase). Additionally some cancers may be linked to protein misfolding. An important example of this is p53-a tumor suppression protein [33].

The cell's ability to recognize, traffic, and metabolize protein aggregates may become defective, impaired, or is overwhelmed. When this occurs a wide variety of diseases are attributed to the presence or accumulation of misfolded protein aggregates. A well known class of protein aggregates diseases are the amyloidoses which will be discussed in further detail below. Instances of non-amyloid protein aggregate pathologies are sickle cell disease, ALS, Alexander disease, and variants of cystic fibrosis.

1.2.3 Aggregation *in vitro*

It is of increasing importance to have a fundamental understanding of protein aggregation not only for understanding human diseases but also for improving the production, purification, and storage of proteins for pharmaceutical therapeutics as well as improving the texture and taste of food.

The explosion of recombinant DNA techniques several decades ago coupled with advances in analytical protein separation technologies have caused a great deal of interest in producing high-grade protein-based pharmaceuticals for a wide-range of clinical and therapeutic applications. However, the risk of aggregation is present at almost every stage of drug development and production [34, 35, 36, 37, 38]. During cell culturing, high levels of expression may cause intracellular aggregates to accumulate. Eventual secretion of the protein into cell culture medium may introduce the protein to conditions which render it unstable. Purification steps such as Protein A chromatography can expose proteins to a regime of unfavorable pH or ionic conditions which could lead to destabilization and aggregation. Storage conditions, be that temperature, freeze-thaw cycles or

the type of container used for storage can have marked influence on the level and amount of aggregation present in a protein pharmaceutical product. It is evident that a fundamental understanding of all the various processes and pathways of protein aggregation is an important task to ensure the continued improvement of the stability and quality of protein-based pharmaceutical products.

Food scientists have been studying protein aggregates for many decades in order to improve the texture and flavor of food. The presence of protein aggregates, and protein aggregate gels in particular, are ubiquitous in food science-from cooked eggs to cheese. The textural mouth-feel of these food products is intrinsically linked with the molecular structure of the protein gels (e.g. fine-stranded vs. particulate gels) and their network connectivity (e.g. the interactions which govern the cross-linking or association of the constituent aggregates). As has been discussed, whey protein aggregate gels can exhibit markedly different rheological and optical properties based upon the conditions in which they are formed which can be very important in producing synthetic food formulations.

1.3 Amyloid Fibrils

As has been mentioned, a large family of diverse human diseases have been associated with filamentous insoluble protein aggregates known as amyloid fibrils. Each disease is related to a particular protein or peptide that undergoes a transition from a native conformation to a highly ordered fibrillar aggregate. These aggregates can then accumulate extracellularly into large plaques, possibly incorporating many other molecular compounds such as sugars or other proteins. Indeed, the term 'amyloid' first derived from the mistaken identification of these plaques as starchy inclusions in *ex vivo* tissue samples.

Some of the most well-known amyloid-related diseases are neurodegenerative disorders such as Alzheimer's disease (aggregation of the peptide fragments amyloid β(1-40) or (1-42)), Parkinsons's disease (α-synuclein), Huntington's disease (Huntingtin), and prion-related spongiform encephalopathies like Creutzfeld-Jakob disease. A number of amyloid fibril related illnesses occur in tissues other than the brain such as Type 2 diabetes (islet amyloid polypeptide), cataract (γ-crystallin), and amyloidosis associated with localized injections, e.g.

Disease	Associated Protein or Peptide
Alzheimer's disease	Amyloid β peptide (1-40) & (1-42)
Parkinson's disease	α-synuclein
Spongiform encephalopathies	Prion proteins
Hungtington's disease	Huntingtin with poly-glutamine expansion
Spinal and bulbar muscular atrophy	Androgen receptor with poly-glutamine expansion
Amyotrophic lateral sclerosis	Superoxide dismutase 1
Spinocerebellar ataxias	Ataxins with poly-glutamine expansion
Dementia with Lewy Bodies	α-synuclein
AL Amyloidosis	Immuglobulin light chains/fragments
AA Amyloidosis	Serum amyloid A protein fragments
Hemodialysis-related amyloidosis	β2-microglobulin
ApoA(I,II,IV) amyloidoses	N-terminal fragments of apoliproprotein A(I,II,IV), respectively
Type 2 diabetes	Islet amyloid polypeptide
Aortic medial amyloidosis	Medin
Cataract	γ-Crystallin
Injection-localized amyloidosis	Insulin

Table 1.1: List of amyloid-related diseases and their associated proteins or peptides

subcutaneous insulin pumps. For a further list of amyloid-related pathologies and their associated proteins and polypeptides see Table 1.1. For a more complete list see [19].

Interestingly, the proteins and peptides associated with these diseases do not share primary amino acid sequences or structural homologies; there is seemingly no apparent universal structural attribute which links these proteins and peptides other than the fact they, under the correct conditions, will undergo the transition to an amyloid fibril state. Indeed, as will be discussed, the key consideration as to whether a protein or peptide is driven to the amyloid state is the conditions under

which fibrils are formed. Both thermodynamic and kinetic factors will determine if a peptide or protein will undergo the amyloid transition.

1.3.1 Structural Schematic of Fibril Formation

The formation of amyloid fibrils can be encapsulated by a generic schematic of the structural species extant along the pathway of assembly. For a typical globular protein, a necessary condition for driving a protein to an amyloid state is to destabilize or partially denature the native conformation. This can be achieved, *in vitro* by alterations in the solution conditions such as temperature or pH, or by additives such as denaturing salts or proteolytic compounds. This destabilization of the native conformation may expose the hydrophobic peptide main chain and residues and seems an important step to inducing amyloid fibril formation [39]. The next step in amyloid fibril formation is the appearance of small, soluble oligomeric species which may appear as amorphous aggregates, micellar structures, or even pore-like rings ([40, 41, 42, 43, 44]). These oligomeric units then go on to assemble into more distinct, filamentous aggregates which are typically on the order of hundreds of nanometers in length and have relatively short persistence lengths ([44]). These 'protofilaments' can then further associate into 'mature' amyloid fibrils which can span many microns in length. See Fig. 1.1 for a pictorial schematic of this process.

1.3.2 Structure & Morphology

Amyloid fibrils, regardless of the initial structural configuration of the native protein or peptide, share many characteristics. Generically, one can utilize three attributes to 'define' any amyloid fibril from other filamentous protein assemblies. Firstly, the archetypal feature of amyloid fibrils is the presence of a cross β-sheet structure as seen from X-ray diffraction [45]. This distinctive pattern is indicative of the protein or polypeptide organized in β-sheets arranged parallel to the fibril axis and their constituent β-strands perpendicular to the fibril axis with spacings of approximately 4 Åand 10 Å, respectively (see Fig. 1.2). Secondly, amyloid fibrils show distinctive optical activity in the presence of histological dyes. Congo red displays 'apple-red' birefringence in the presence of amyloid fibrils. Another dye, Thioflavin T, is a benzothiazole salt which displays

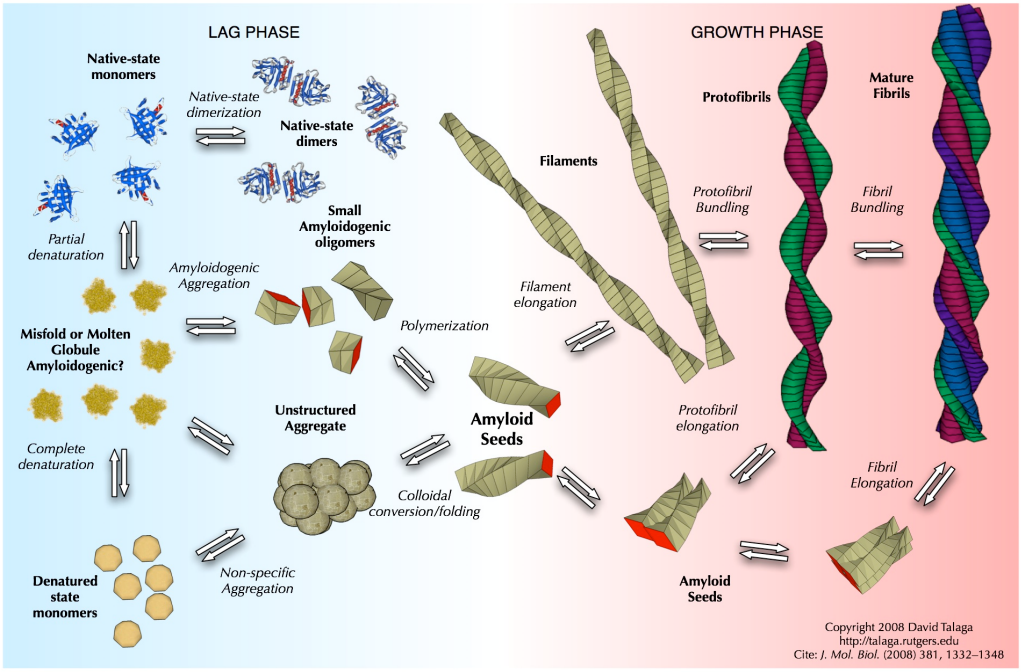

Figure 1.1: **Generic picture of amyloid fibril assembly.** For structured proteins or peptides, it is commonly accepted a destabilization of the native structure must be achieved before amyloidogenic aggregation can proceed. From a partially denatured monomer unit, amyloid or amyloid-competent oligomers form until amyloid 'seeds' or nuclei are formed. From this point, amyloid filaments may grow by addition to the ends of the growing filaments. These filaments can then associate into protofibrils and eventually mature amyloid fibrils. Image taken from [1].

characteristic fluorescence enhancement when bound to the structure of amyloid fibrils ([46, 47, 48]). Thirdly, amyloid fibrils, as seen by atomic force microscopy (AFM) or transmission electron microscopy (TEM), exhibit typically filamentous morphologies with typical diameters of 2 to 20 nm and lengths up to several microns ([49]).

Advances have been made to elucidate the organization of amyloid fibrils at a molecular level using high resolution techniques such as solid-state NMR spectroscopy, X-ray crystallography, and cryoelectron microscopy. The picture that emerges is that peptide backbone, mainly via hydrogen bonding, determine the overall structural framework at the core of amyloid fibrils. Indeed, the

ubiquity of cross-β sheet structure in amyloid fibrils is highly indicative that the peptide backbone is the major player in determining the generic structure of amyloid fibrils.

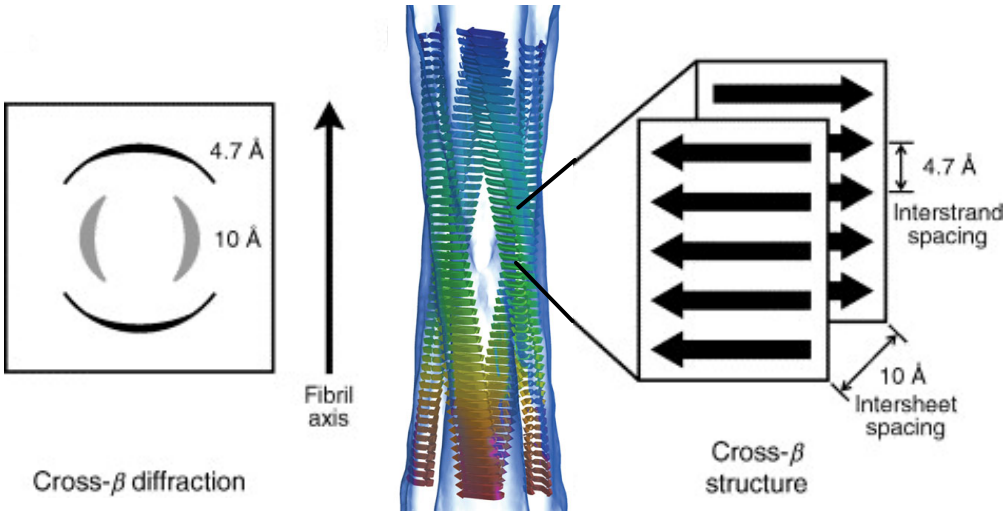

Figure 1.2: **Generic structure of amyloid fibrils.** Amyloid fibrils are typified by the presence of cross β-sheet structure as revealed by x-ray diffraction experiments. The schematic shows a meridional x-ray diffraction pattern indicating a spacing of $\sim$ 10 Å. A second equatorial pattern shows a spacing of 4.7 Å. These spacing correspond to the inter-sheet and inter-strand spacing within an amyloid fibril. Images adapted from [2].

At the next hierarchical level of structure, these techniques have shown there exists a degree of polymorphism between fibrils created from different peptides or proteins in terms of how the β-strands are oriented, the number of β-sheets in a protofibril, how many protofibrils constitute a mature amyloid fibril [44, 50, 2, 51]. These differences may be attributed to how the side chains within the fibril interact with one another sterically and electrostatically and also with the surrounding solvent.

This molecular-level polymorphism, driven by side-chain interactions can generate heterogeneous morphologies of mature fibrils, even those formed from the same peptide or protein ([2, 51, 52]). One may find, under the correct conditions, long, straight fibrils ([53, 54]), wide tapes ([55]), twisted ribbons

([53, 56]), or closed loops ([56, 57]). Such differences in morphology may be linked back to differences in the molecular structure of these fibrils caused by different conditions in which the fibrils form ([58]). Moreover, it has been shown that in many cases, fibrils formed from seeds inherit the structural characteristics of the parent seeds. It is interesting to consider that, regardless of the amino acid sequence of a peptide or protein, there is a spectrum of structures available in the free energy landscape and the existence of one morphology over another is controlled thermodynamically and kinetically, based upon the conditions of formation.

Many researchers have made an effort to identify possible indicators in the amino acid sequence of peptides and proteins that may point toward a propensity to form amyloid fibrils and several criteria have been identified. First, hydrophobicity seems to be an important factor. Studies have shown making amino acid substitutions with higher (or lower) hydrophobicity can increase (or decrease) the propensity of fibril formation ([59, 60]). Indeed, it has been demonstrated that in natural protein sequences, there is an absence of clusters of hydrophobic residues within a sequence [61, 62]. It seems evolution has selected amino acids sequences to avoid three or more consecutive hydrophobic residues. Second, proteins or peptides with high electrical charges will not easily self-associate due to electrostatic repulsion, thus decreasing the propensity to aggregate([63]. Thirdly, it has been shown that sequences that promote the formation of β-sheets over α-helices are lower than expected, indicating evolution has selected against sequences that form β-sheets ([64]).

1.3.3 Universality of the amyloid fibril structure

Despite these insights into how amino acid residues can influence not only the propensity to form amyloid fibrils but also their morphology, what has clearly emerged is that the critical interactions that drive the assembly of peptides and proteins into the amyloid state occurs via the peptide backbone. The observation that there is no characteristic amino acid sequence or structural homology which is universally indicative of amyloid forming proteins and peptides can be understood through the fact that amyloid assembly is most critically dependent on the backbone association into β-sheet architecture. It is conceivable that amyloid fibril structure represents a global minimum in the free energy landscape of

peptides and proteins ([65]). Therefore, any peptide or protein theoretically can be driven to the amyloid state. As has been noted, what ultimately determines if and when this occurs is based on thermodynamic considerations and the kinetic availability of the amyloid state which is contingent on the solution conditions during formation.

1.3.4 Functional Amyloid Used by Living Organisms

It has been found that some living systems have utilized amyloid fibrils in a functional and beneficial way. Several species of bacteria have developed highly regulated biomolecular processes to assemble functional amyloid to aid in biofilm formation and spore dispersal [66, 67, 68]. There have even been examples found in mammalian systems which have developed highly specific and regulated use of amyloid-like structures. A domain of the membrane protein Pmel17 forms an amyloid like structure within the compartmentalized melanosome to template the formation of melanin granules [69].

1.3.5 Amyloid Fibrils as Biomaterial

Amyloid fibrils possess several characteristics which make them attractive as a class of biomaterials. Firstly, amyloid fibrils self-assemble. Self-assembly is the process by which molecular objects, via cooperative molecular recognition, form regularized supramolecular structures without external input. Self-assembly is ubiquitous in biology – from micellization to protein folding. This phenomenon is an indispensable tool in any bottom-up approach to constructing nanomaterials.

Secondly, they possess structural properties unique to many biological materials. A range of experimental techniques have demonstrated amyloid fibrils possess a Youngs Modulus around the GPa scale and a strength comparable to steel [70]. The structural properties of any material gives indirect indications on the underlying intermolecular forces which collectively give rise to the bulk attributes of a material. The distinguishing feature of amyloid and amyloid-like fibrils, according to high resolution data gathered from x-ray diffraction and NMR experiments, is the presence of extensive cross β-sheet structure within the core of the fibril. This extensive β-sheet structure forms an expansive hydrogen-bonding network involving the polypeptide backbone which spans the length of

the fibril. It is this cooperative intermolecular hydrogen bonding network which confers the stability and unique material properties to amyloid fibrils [71]. While a complete structure of an amyloid or amyloid-like fibril has yet to be determined this common structural motif is always present, regardless of the initial protein or polypeptide conformation. Indeed, amyloid and amyloid-like fibrils are known to form from large globular proteins with diverse secondary and tertiary structure down to short dipeptides and protein fragments. Ultimately, amyloid fibrils, regardless of the initial structure, generally exhibit typical dimensions of 5-20 nm in diameter and extend to microns in length. Furthermore, the fact that amyloid fibrils exhibit similar mechanical properties to silk, which is known to be one of the strongest and most rigid proteinacious materials in nature, should not be surprising; silk and amyloid fibrils both possess the common characteristic of expansive hydrogen-bonded β-sheet networks.

Thirdly, since amyloid fibrils can be constructed from a diverse range of proteins and polypeptides, the conditions by which fibrils form are also expansive. The one apparent key environmental characteristic to form amyloid fibrils is the existence of partially denaturing conditions which destabilizes the initial conformation. This flexibility in experimental conditions lifts potential restrictions in applying the assembly of amyloid fibrils to nanobiotechnological applications.

Since amyloid fibrils are constructed from amino acid units the power and relative ease of genetic modification and manipulation of bacterial expression systems is at the disposal of anyone designing novel amyloid biomaterials [72]. These well understood techniques permit the introduction of point mutations in the amino acid sequence which opens the possibility to chemically functionalize the surface of a fibril or changing its electrostatic characteristics [73, 74, 75]. Furthermore, given the enormous combinatoric sequence space available for creating de novo peptide and protein sequences, it should be possible to design amyloidogenic materials with unique environmental responsiveness. There has already been some success in utilizing amyloid fibrils in biomaterial applications from templating nanowires [76, 77, 78], hydrogels for tissue engineering [79, 80, 81, 81, 82, 66] and drug delivery systems [81, 83, 84, 85], to light harvesting materials [86, 87, 88]. The potential ease of modification, hybridization and functionalization vastly enlarges the material applications, capabilities, and

potential usefulness of amyloid as a biomaterial.

1.4 Kinetic Models of Amyloid Fibril Formation

It is clear a full understanding of the complete process of amyloid fibril formation requires a detailed understanding of the species extant at each phase of assembly. In particular, an elucidation of the molecular contacts each species makes to move onto the next step of fibril assembly and the thermodynamic and kinetics of each of these transitions is of great importance. However, at a more coarse-grained level of detail, the kinetics of amyloid fibril assembly can be simply characterized by a few generic features: a lag phase in which no fibrils have been formed (within the limits of experimental detection), this is followed by a rapid proliferation and elongation of fibrils after the quiescent lag phase and a final steady-state phase where growth has come to completion [39, 89, 90, 91, 92]. Additionally, the lag phase can be eliminated by the addition of pre-formed fibrillar seeds and there is a critical initial concentration below which fibrils will not form [49].

There are many ways in which the kinetics of amyloid assembly can be monitored. ThT fluorescence is a powerful and common method to measure the assembly process since ThT preferentially binds to the characteristic cross β-sheet structure of amyloid fibrils [39, 93]. Although the exact mechanism of ThT binding to amyloid fibrils is not completely understood [46, 47, 48] it has been established that it has a negligible effect on the self-assembly process [39]. One must take care, however, when utilizing ThT as a reporter of the kinetics due to potential non-linear scaling with concentration [10]. Additionally, ANS fluorescence has shown that fibril formation is indeed preceded by partial unfolding and the exposure of the interior hydrophobic core for structured and globular proteins [39]. Dynamic light scattering (DLS) has proven useful not only for monitoring the kinetics of fibril growth but also as a means to extract the size (hydrodynamic radii) of molecular species during the time course of the experiment [94, 95, 96, 97, 98]. DLS, while useful to study early events and species in the fibril formation process, it is unable to provide detailed information regarding aggregate sizes once large species are present which overwhelm the scattering signal. Another useful means of characterizing the oligomeric content during the time course of assembly is by mass spectrometry [99, 100]. Mass

spectrometry data suffers from the drawback that the monomer or oligomeric species being studied are removed from the solution environment and subject to harsh gas-phase conditions which can influence their structure. Circular dichroism (CD) and Fourier Transform Infrared Spectroscopy (FTIR) provide a means to monitor the changes to secondary structure of molecular constituents during aggregation [101, 102, 98, 54, 103]. Atomic force microscopy (AFM) is a powerful tool to gain insight into the oligomeric species present and their evolution during self-assembly [104, 105, 106, 52, 107]. Along with AFM, electron microscopy (EM) techniques have also shed light on protofibril structure and a means to visualize the wide array of fibril morphologies [2, 51, 90, 108, 109]. Other techniques such as static light scattering, small angle X-ray scattering [110, 90], neutron scattering [111, 112], and solid state NMR [113, 114] can be useful in elucidating the kinetics, with molecular resolution, of fibril formation. No single technique is sufficient to elucidate the processes and mechanisms of self-assembly; a program of study which includes some of these can provide complimentary and corroborating information which ultimately can allow one to begin to understand the system being studied.

1.4.1 Classical Nucleation Model

The features of amyloid fibril assembly, as listed above, bear the hallmarks of a nucleation-dependent growth mechanism: the lag phase is the time it takes to form a critical nucleus, from which rapid growth by addition can then follow. The final equilibrium is reached once the concentration of protein drops to the critical concentration. Indeed, there are many protein aggregation processes which are governed by nucleation-dependent kinetics. In this section, I will review the standard classical nucleation models for protein aggregation, and in particular apply these ideas to understand the observed kinetics of amyloid fibril formation.

1.4.1.1 Homogeneous Nucleation

One can define in a formal way, the formation of linear aggregates by end-to-end association,

$$X + X_i \rightleftharpoons X_{i+1} \tag{1.1}$$

where X is a monomer unit and X_i is an aggregate composed of i-monomer units. This reaction is of course characterized by the polymerization and depolymerization rate constants, k_i^+, k_i^-, respectively, for all species i. In many cases, the initial step(s) in this reaction have rates much smaller compared to later ones even if the conditions for nucleated polymerization are favourable. This is due to the fact that in order for polymerization to proceed, a thermodynamically unfavourable nucleus must be formed. To make this more concrete, one can simplify this formal scheme by considering that the initial steps of this reaction take place close to thermodynamic equilibrium and that these initial steps, up until a certain critical size of aggregate, n^*, is thermodynamically unfavourable,

$$X + X_i \longleftarrow X_{i+1}, i < n^* \tag{1.2}$$

$$X + X_i \longrightarrow X_{i+1}, i > n^* \tag{1.3}$$

The probability of finding a particular species X_i, at equilibrium, is of course related to the Gibbs free energy,

$$[X_i] = [X_0] \exp(-\Delta G(i)/RT) \tag{1.4}$$

where X_0 is some standard state concentration, often taken as the initial concentration. With this in mind, one can write down, at equilibrium,

$$k_i^+ c[X_i] = k_{i+1}^-[X_{i+1}] \tag{1.5}$$

which can be re-written as a ration of the relative concentration of species i and $i+1$, $[X_{i+1}]/[X_i] = ck_i^+/k_{i+1}^-$. This ratio can be cast in terms of the ratio of Gibbs free energies such that,

$$\frac{[A_{i+1}]}{[A_i]} = \exp(-\frac{d\Delta G}{di}/RT) \tag{1.6}$$

Therefore, one can understand the transition in the rate constants in Equations 1.2 & 1.3 as the change in the slope of the free energy barrier. For a sufficiently steep barrier, there exists essentially one state, populated by a very small number of species n^* such that it becomes the rate-limiting step in the reaction. This species n^* is defined as the critical nucleus; the kinetics of aggregation are therefore determined by the population and the growth rate of the nuclei [115]. The waiting time for nucleation formation can be bypassed by 'seeding' the system with pre-formed nuclei. By doing so, the growth kinetics may bypass the lag phase and growth may immediately occur. Such a kinetic scenario is referred to as a downhill polymerization process when the forward rate of elongation is essentially irreversible and has a first-order dependence on concentration [115].

In this simple scheme what do the kinetics look like? One can write down equations for the polymerization rates involving growth rates of nucleus, the rate of monomer addition to polymers, etc. which can be solved by a number of methods [116, 117]. What emerges from this formalism are some key features. Firstly, in this simple linear aggregation picture with a homogeneous nucleation step, the early rate of growth is proportional to t^2 and results in a weak delay to the start of polymerization. Secondly, the characteristic time of the aggregation reaction has a very specific scaling relation with the initial protein concentration. Oosawa showed that the characteristic time, or the lag time, τ scales with an exponent γ as $\tau \sim c^{-\gamma} = c^{-(n^*+1)/2}$. The important point is that the exponent γ is always greater than unity. It is therefore very instructive to study aggregation kinetics as a function of initial protein concentration. As an application of this theory, Oosawa demonstrated this homogeneous nucleation scheme describes well the polymerization of F-actin [116]. It should be noted that the polymerization of F-actin into linear aggregates is an example of 'native' self-assembly.

1.4.1.2 Secondary Pathways of Aggregation

While the previous theory describes actin polymerization well, there are cases of protein aggregation which show much more pronounced lag times and exponential growth after the lag phase. One example of a protein aggregating system with pronounced lag times followed by exponential growth is the polymerization of sickle cell haemoglobin [118]. There are several ways of achieving this type of behavior: fragmentation, heterogeneous nucleation, or branching. Fragmentation is merely the breaking up of aggregates into smaller pieces; this allows for new ends to be created resulting in an autocatalytic and exponential increase in the growth of aggregates. Heterogeneous nucleation is the nucleation of a new aggregate on the surface of an already formed aggregate. Branching and heterogeneous nucleation on the face of it seem identical–however they are thermodynamically distinct. There is no free energy barrier governing the rate of branching; heterogeneous nucleation is dependent on a free energy barrier (which is often lower than the primary nucleation barrier) that must be traversed before additional growth may occur.

These features may be incorporated into a similar formalism as discussed above, kinetic equations written down and solved [115]. Several salient features again emerge. Firstly, the initial growth rate is exponential. Secondly, the lag time, which is governed by the primary nucleation process, will still have the same concentration scaling as described above. Thirdly, if one observes a pronounced lag time in the kinetics of aggregation it does *not* necessarily imply the presence of a nucleation-dependent aggregation process. The pronounced lag time present in secondary processes is a result of the exponential character of the solution to the kinetic equations and *not* nucleation. As has been stated, in order to understand the origin and details of the kinetics of an aggregation process, one must study the concentration dependence of the various kinetic parameters one uses to describe the system. This type of formalism with a secondary pathway was extremely successful in explaining the kinetics of sickle cell haemoglobin via a heterogeneous nucleation mechanism [23, 118].

1.4.2 Fragmentation driven self-assembly of amyloid fibrils

In the preceding sections, the kinetic features of aggregation and assembly have been predominately dictated by primary nucleation events. For instance, the scaling behavior of the lag time is determined by the details of rate of primary nucleation and the size of the nucleating species. A new analytical model has been developed which combines primary and secondary nucleation events, coupled to linear growth, that interestingly demonstrates that the kinetics of growth can be dominated by the secondary rather than primary processes [119, 120, 121, 122]. In their approach they write down a Master equation detailing the coarse-grained details of the microscopic processes that contribute to fibril assembly,

$$\begin{aligned} \frac{\partial f(t,j)}{\partial t} &= 2m(t)k_+ f(t,j-1) \\ &- 2m(t)k_+ f(t,j) - k_-(j-1)f(t,j) \\ &+ 2k_- \sum_{i=j+1}^{\infty} f(t,i) + k_n m(t)^{n_c}\delta_{j,n_c} \end{aligned} \tag{1.7}$$

where m(t) is the concentration of monomers, f(t,j) is the concentration of fibrils of length j, k_+ the elongation rate by monomer addition, k_- the fragmentation rate, n_c is the critical nucleus of size n, and k_n is the nucleation rate. The first term represents the increase in the number of fibrils of length j by the addition of a monomer to either end of a fibril, the second term is the decrease of the number of fibrils of length j growing to j+1, the third term describes the breaking of a fibril at any one of the j-1 possible sites along the its length. The summation term accounts for the fact that there are two sites in any fibril which, after fragmentation, will result in a fibril of length j and the final term is the rate of nucleation of nuclei of size n.

From these non-linear and infinitely coupled kinetic equations, they are able to write down the principal moments of $f(t,j)$ in closed form to a very good approximation. What emerges is a relatively straight-forward picture: the maximal rate of fibril assembly depends only on one kinetic parameter, $\kappa = \sqrt{2m_{tot}k_+k_-}$ (this is the rate of multiplication of the fibril population and m_{tot} is the total protein concentration). No other kinetic parameters like k_n are

important to describe the details of the assembly kinetics. Moreover, they can write down an expression for the lag time as a function of this parameter κ,

$$\tau_{lag} = [log(1/C_+) - e + 1]\kappa^{-1} \tag{1.8}$$

where the C_+ term contains details about rate of nucleation. Interestingly, the lag time is *not* dominated by the formation of critical nuclei–this only appears as a small logarithmic term in the above equation. The important conclusion to be drawn is that when the ratio $k_-/m_{tot}k_n n^c \gg 1$, the secondary process active in the kinetics, in this case fragmentation, will determine the details of the kinetics, such as the lag time.

Since the kinetics are essentially governed by this single parameter κ one can obtain a scaling law for the lag time with concentration similar to that found in the classical nucleation picture. As we have seen $\tau_{lag} \sim \kappa^{-1} \approx m_{tot}^{-\gamma}$, where γ in this instance is 0.5. This is in contrast to the result found in the classical nucleation model, where γ is always greater than unity. Therefore, by looking at the concentration dependence of the lag time, one can determine if the kinetics are dominated by a secondary process such as fragmentation.

1.4.3 Further details on the lag time of amyloid fibril assembly

We have seen the lag time is an important kinetic parameter for the formation of amyloid fibrils-one can determine possibly the size of the critical nucleus or the kinds of processes active in fibril assembly based upon the way the lag time scales with concentration. Furthermore, recent discoveries have shown that the most toxic species in amyloid fibril formation are not the fibrils themselves but the prefibrillar aggregates [40] often present in the early times of assembly. Hence, a complete picture of the lag phase of amyloid fibril assembly is an important task to resolve.

Fändrich has shown, by studying the lag times and growth constants of a variety of amyloidogenic protein systems, there exists a universal underlying mechanism in their nucleation behavior [92]. The lag time is also an indicator for the propensity of an amyloidogenic protein system to form fibrils. The lag

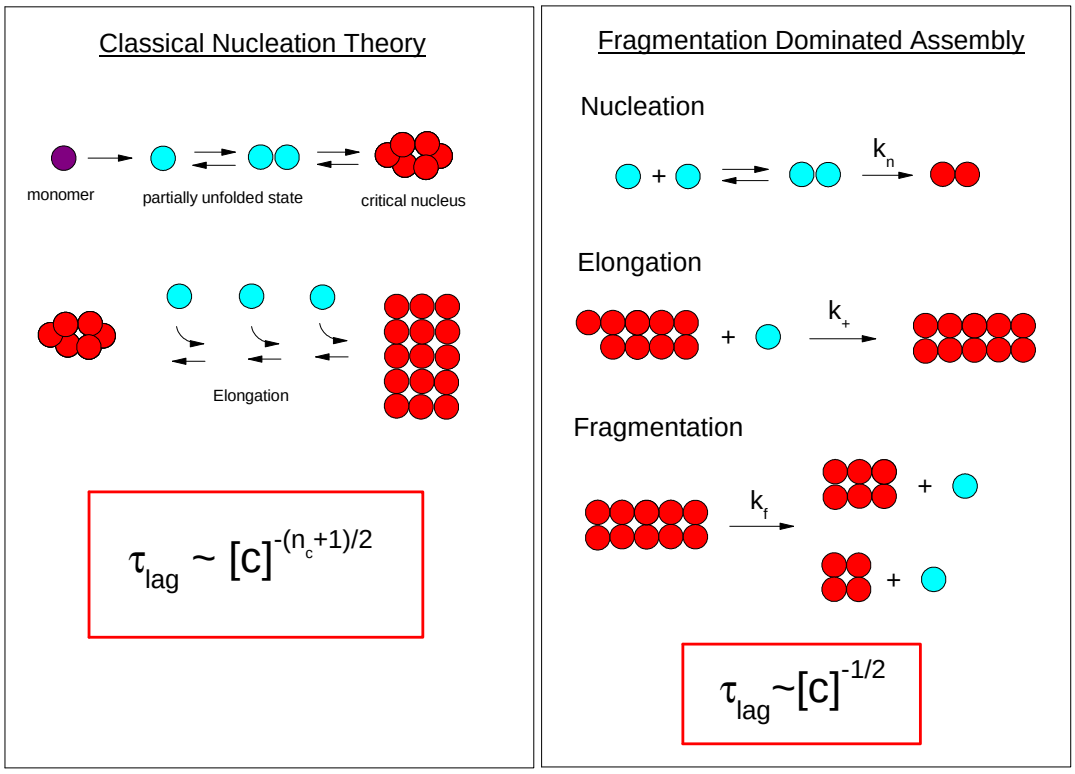

Figure 1.3: **Schematic of kinetic models.** In the classical nucleation picture of amyloid fibril assembly, a native monomer unit becomes partially destabilized due to environmental conditions. These destabilized units can aggregate, eventually forming a critical nucleus of some size n_c that can then continue to grow into fibrils. Such nucleated growth is typified by the lag time scaling with the protein concentration as $\tau_{lag} \sim c^{-(n_c+1)/2}$. In a model where nucleation is still present, but the kinetics are dominated by fibril fragmentation as different scaling behavior exists quite distinct from the classical model: $\tau_{lag} \sim c^{-1/2}$. Since the scaling exponent in the classical nucleation picture is always greater than one, the predominant mechanisms which determine the kinetic features of fibril assembly may be reliably distinguished.

time is extremely sensitive to solution environment such as pH, salt content, temperature, agitation and protein concentration[39, 123, 124]. By studying the lag time, information can be gained about the factors that influence fibril formation.

An open question concerning the lag phase is whether it is a deterministic or stochastic process. Hofricther observed that the onset time of sickle cell haemoglobin polymerization fluctuated and a distribution of lag times was generated. This distribution was described by the nucleation event, i.e. crossing

the free energy barrier, being a random walk. He then used this distribution of lag times to deduce the rate of nucleation [23]. Szabo went on to a provide a detailed mathematical description of the lag time distribution to determine accurately the nucleation rates of sickle cell haemoglobin polymerization [125].

Hortchansky *et al.* observed variability in the lag time of Aβ(1-40) fibril formation, as a function of concentration, from identical samples (splitting of a stock solution into individual aliquots) and that the lag time variation is inversely correlated to the concentration. [123]. They obtained a Gaussian lag time distribution, which they claim is evidence for stochastic nucleation. Their distributions are constructed of approximately 20 replicates per concentration so their statistical analysis is not completely conclusive. They also posit that this macroscopic heterogeneity evidenced by the distribution of lag times may be connected to the heterogeneity known to exist at the microscopic level. Therefore the stochasticity of the nucleation event may point to the appearance of varied morphologies of oligomers, protofibrils, and mature fibrils.

Foderá *et al.* also observe similar variation in the lag time, in this case during the fibrillation of human insulin [10]. They undertake a statistical study of these lag time variations and find that the spread of the distribution is dependent on the concentration. As in [123] the spread is much greater at lower concentrations. They make the analogy between insulin and sickle haemoglobin polymerization: the stochastic initiation of a single nucleation event followed by secondary, heterogeneous growth of insulin fibrils. Indeed, it has been shown that insulin exhibits heterogeneous nucleation akin to sickle haemoglobin polymerization [93]. Their statistics are only based on an approximate sample size of 50 kinetic traces. As in [123] the statistics, albeit pointing toward stochasticity, are somewhat inconclusive. The precise description of this lag time distribution has yet to be described.

Recently, however Hellstrand *et al.* have claimed that the stochasticity observed in these and many other experiments are an artifact of imprecise experimental protocols and a wide variation in experimental conditions [126]. They look at the Aβ(1-42) aggregation monitored by ThT fluorescence assays and enzyme-linked immunosorbent assays (ELISA). By having a strict protocol for production and filtration of the peptide and the use of PEG-coated, non-binding assay plate, they are able to produce Aβ kinetics with great reproducibility.

Previous lag time variation was on the order of hours. They obtain variations on the order of minutes. They also have improved statistics by looking at 672 kinetic traces over varying concentrations. They conclude that in fact there is no stochasticity in the nucleation event; this was merely a by-product of experimental imprecision. They fit the lag time versus concentration by a power law with an exponent of -1.5. They find that, akin to Flyvbjerg [117], the critical nucleus is of size 2. This conclusion is drawn from the assumption that there are no secondary pathways.

Taken together, one must conclude that the attribution of variability in the lag time, almost certainly is not due to nucleation, as has been common in the literature, but to imprecise experimental protocols and techniques. For amyloid assembly that is dominated by fragmentation, the explanation that the variability is due to nucleation are incompatible with one another. Experimental volumes in almost all studies exploring the kinetics of fibril assembly ensure that nucleation events occurs rapidly and the subsequent features of the kinetics are dictated by the elongation and fragmentary processes of assembly. Such variation in lag times can only be understood in terms of nucleation in the limit of small system volumes (c.f. [127]). Therefore, the standard line in the literature concerning nucleation and the lag time must be re-evaluated for protein systems in which it can be shown that the kinetics are best described by a fragmentation-type model.

However, since the stochastic nature of nucleation determines amyloid growth lag times in the small volume limit it may be possible to begin to understand why, for instance, there may be variation in the onset of amyloidogenic diseases in genetically identical individuals. An understanding of these variations can help aid in the potential development of inhibitory treatments targeting the early stage pre-fibrillar aggregates present in the lag time.

1.5 Outline Of Thesis

This thesis will study the formation of three different protein systems that form fibrillar aggregates. First, in Chapter 3 I study mechanism of insulin amyloid fibril formation and find a structural transition which has many interesting ramifications concerning the growth kinetics, fibril structure, rheology, and the presence of potentially harmful cytotoxic aggregate species. This transition will

be further studied using small angle neutron scattering in Chapter 4. Second, I explore the mechanism of ovalbumin fibril formation in Chapter 5. I find that the ovalbumin growth kinetics contains several interesting features, some of which may be understood in terms of the morphological characteristics of the fibrillar aggregates. Finally, in Chapter 6 I investigate the curious case of ocr gelation. Under the conditions in which ocr forms fibrillar aggregates, most proteins would form amorphous aggregates and precipitate from solution. However, this protein forms a translucent fine-stranded gel at very high salt concentration. I study the structure of gels formed from mutant variants of ocr to better understand the underlying mechanism of this unusual instance of fibrillar protein aggregation.

Chapter 2

Materials & Methods

2.1 Protein Systems Studied

2.1.1 Insulin

Insulin is a 51 amino acid (M_w=5808 and 5733 Da for human and bovine insulin, respectively) peptide hormone created in β-cells within the pancreas that is involved in the regulation of fat and carbohydrate metabolism. Insulin is comprised of two chains: the A and B chains which are linked by two disulphide bonds. There is a further third internal disulphide bond within the A chain (see Fig. 2.1). The secondary structure of insulin consists primarily of α-helices while the disulphide bonds stabilize the tertiary structure. The monomeric unit of insulin acts as the biologically active form [128]. The monomer contains two non-polar surfaces; one is mainly comprised of aromatic residues while the other is more extensive and flexible [129]. These surfaces are reactive and act as the sites for receptor interactions [130]. However, over a large pH range (2-8) insulin exists in a dimer state, burying the aromatic non-polar surface. Dimeric insulin undergoes further assembly into a hexameric unit that is coordinated by the presence of 2 to 4 Zn^{2+} ions [129]. Low pH, elevated temperatures, and agitation can induce insulin to be in the reactive monomeric state [131].

This monomeric state appears to be an important ingredient in driving insulin to form amyloid fibrils [39, 132, 101]. Indeed, insulin has a long history as a test system to study the formation of amyloid fibrils going back to the 1940s and 1950s and the work of D.F. Waugh [133, 134]. Generally, insulin forms amyloid

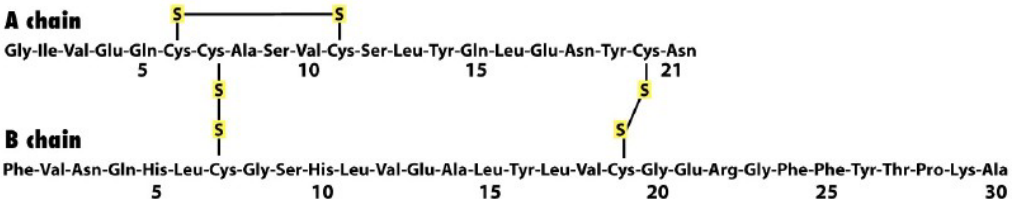

Figure 2.1: **Schematic of the amino acid structure of bovine insulin.** Bovine insulin consists of 51 amino acids which are subdivided into two units: the A and B chains. There are two inter-chain disulphide bonds linking the A and B chains; there is also an interchain disulphide bond within the A chain. Taken from [3]

fibrils at low pH (pH < 2) and high temperatures ($>37°$C). The rate at which insulin forms amyloid fibrils is contingent upon many parameters from the kind of acid used, amount and kind of salt present, temperature, degree of agitation, and the presence of surfaces [39, 135]. Insulin has been a well-studied amyloidogenic system due to its relatively inexpensive cost and ease in which it will form amyloid fibrils [90, 54, 10, 136, 137, 138, 105, 2, 139, 104, 140, 110].

In this thesis, bovine insulin was used to study the kinetics of amyloid fibril formation in a variety of conditions and concentrations. Insulin is not only an attractive test system for the reasons mentioned above, but also has direct consequences for the pharmaceutical industry. For example, the accumulation of insulin amyloid deposits in subcutaneous insulin injection systems for Type-1 diabetes patients is an ongoing issue [141] as well as prolonging the shelf-life of pharmaceutical suspensions of insulin. Therefore studying insulin amyloid fibril formation can have a broader impact than just being a 'cheap standard protein' for studying amyloid fibril formation.

2.1.2 Ovalbumin

Ovalbumin is the primary protein constituent in hen egg white. Its ubiquity and thorough structural characterization has made it a useful test protein system. Interestingly, its function is still not understood [142]. Ovalbumin consists of

386 amino acids ($M_w \sim 48$ kDa) that, under native conditions, is comprised of α-helix, β-sheet, and random coil in a compact globular configuration containing one intra-protein disulphide bond (see Fig. 2.2).

Several papers have shown that ovalbumin will form amyloid fibrils by reducing the internal disulphide bond and/or incubating ovalbumin in denaturing conditions [143, 144, 145]. I sought to study the formation of amyloid fibrils from ovalbumin due to its unusual kinetic behavior as well as its copious availability. In this thesis, ovalbumin was induced to form amyloid fibrils by the adding a reducing agent and heating (§2.2.1.3).

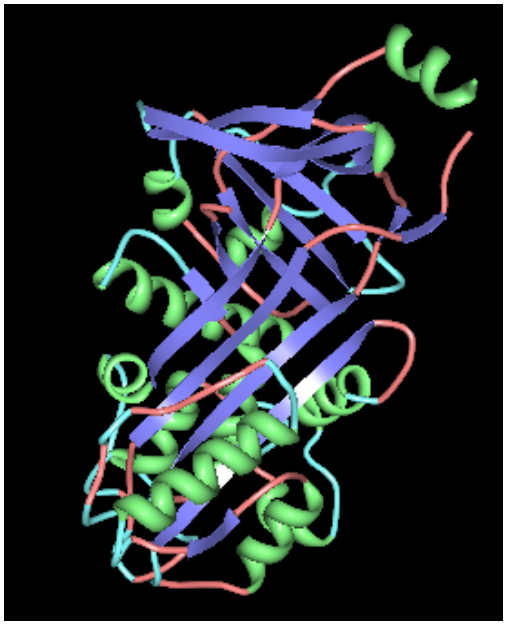

Figure 2.2: **Ovalbumin structure.** Displayed is an ovalbumin monomer unit. Ovalbumin contains substantial β-sheet (dark blue) and α-helix (green) secondary structure. A member of the serpin family of proteins, it is the most abundant protein in a hen egg white. Taken from PDB 1OVA.

2.1.3 ocr

Many microbial organisms have developed defense mechanisms to protect themselves from viral infections. During a viral infection, the virus will attach itself to the host microbe and inject its DNA. However, the microbe has a means to distinguish its own DNA from the foreign viral DNA. An enzyme known as a

methyltransferase (MTase), which methylates unique sequences within the host genome combined with an enzyme called a restriction endonuclease (REase), can distinguish the methylated host DNA from the unmethylated DNA of the virus. The REase will then destroy the viral DNA thus preventing infection. This process is called the restriction modification system (R/M). An additional enzyme is present that determines the specificity of the R/M process. This is known as a type I restriction/modification complex. Viruses known as bacteriophages have evolved unique methods of circumventing this R/M system such as methylating their own DNA. However, bacteriophage T7 has evolved a unique strategy: a DNA mimic protein [146].

2.1.3.1 What is Ocr?

When bacteriophage T7 first infects *Eschericia coli*, the first protein that is expressed is ocr (Overcome Classical Restriction). Ocr is a protein consisting of 116 amino acids in monomer form with a molecular weight of 13678 Daltons (see Fig. 2.3 for the amino acid sequence of wild-type ocr). Importantly, ocr has a very high number of aspartic and glutamic acids arranged on its surface giving it a very high negative charge. At pH 8.0, for instance, the ocr dimer (ocr in solution is dimeric) has a charge of -56e. Ocr contains all the natural amino acids apart from cysteine. The exclusion of cysteine precludes the formation of inter- or intra-protein disulphide bridges. The tertiary structure is comprised of four α-helices per monomer. There are no β-sheets present in ocr.

```
1   MAMSNMTYNNVFDHAYEMLK
21  ENIRYDDIRDTDDLHDAIHM
41  AADNAVPHYYADIFSVMASE
61  GIDLEFEDSGLMPDTKDVIR
81  ILQAIYEQLTIDLWEDAED
101 LLNEYLEEVEEYEEDEE
```

Figure 2.3: **Ocr amino acid sequence** Wild-type ocr is comprised of 116 amino acids. There is a large number of aspartates and glutamates which contributes to the very high surface charge of ocr.

Ocr is expressed first since it is an inhibitor of *E. coli's* R/M system. Once

expressed, it will bind to an REase (for *E. coli* the type I enzymes are called EcoKI and EcoBI), preventing the enzyme from recognizing and destroying the viral DNA. The viral DNA is able to freely transcribe and the virus replicates. Ocr's inhibition of the R/M system is achieved by mimicking the structural and electrostatic properties of DNA. The carboxyl groups of ocr simulate the distribution of phosphate groups of B-DNA [147]. Furthermore, when EcoKI binds to DNA it causes a bend to accommodate the target recognition domains.

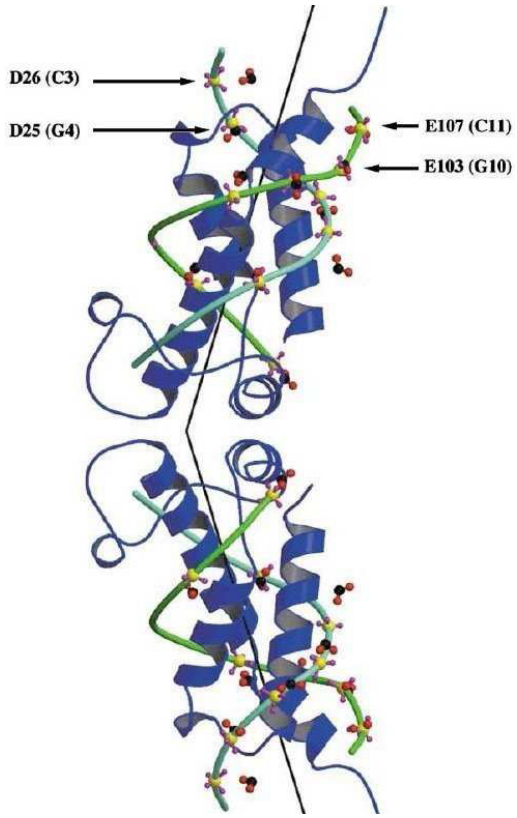

Figure 2.4: **The Ocr dimer dimics the charge and structure of bent DNA'** An ocr dimer (blue) with bent double stranded DNA (green) overlaid. Note the similarity in location of the charge groups for ocr and DNA. DNA phosphate groups are yellow and purple. The ocr carboxyl groups are red and black. The bend angle is 33.6°. (Source: [4])

Ocr exists naturally as a dimer; dimeric ocr has a natural bend very similar to a bent DNA molecule bound to EcoKI [148]. Fig. 2.4 shows a dimer of ocr with bent double stranded DNA laid over top. One can see that the arrangement of charge groups on ocr closely resembles that of the DNA. In fact, ocr binding

affinity to type I enzymes EckoKI and EcoBI is 50 times greater than that of DNA. Ocr is therefore very effective in competing with DNA in binding to type I enzymes; so much so that ocr can even cause bound enzymes to disassociate from DNA [147]. Indeed, ocr is so effective because it will bind virtually all R/M enzymes regardless of sequence specificity by virtue of its shape and charge characteristics [148, 149].

2.2 Experimental Methods

2.2.1 Thioflavin T & Amyloid Fibrils

Thioflavin T (ThT) is a dye which preferentially binds to the structure of amyloid fibrils. Free ThT in solution weakly fluoresces with $\lambda_{ex} = 350nm$ and $\lambda_{em} = 440nm$. It was observed in the late 1950s that ThT exhibits characteristic fluorescence enhancement when bound to amyloid fibrils at $\lambda_{em} = 482nm$ when excited at $\lambda_{ex} = 450nm$ [150]. Unlike other amyloid specific dyes, such as Congo Red, it has been shown that ThT does not interfere with the assembly process or significantly perturb the structure of the fibrils [151, 48]. Because of this, ThT is particularly useful in monitoring the kinetics of amyloid formation and growth. Despite the utility of ThT, the details of the binding mode to amyloid fibrils are still not well understood. Krebs *et al.* demonstrated that the orientation of binding as well as the steric interaction between ThT and the side chains of the fibrils contributes strongly to the fluorescence enhancement observed [46]. They propose ThT binds to 'channels' that run along the β-sheet structure of amyloid fibrils. Similarly, Groenning *et al.* have shown that the characteristic ThT fluorescence is induced by ThT binding within cavities which can accommodate one or more ThT molecules. They posit that the presence of multiple ThT molecules within a cavity can induce excimer formation which may explain the enhancement of fluorescence often observed [47].

2.2.1.1 Thioflavin T Fluorescence Assays: Experimental Procedures

Thioflavin T fluorescence assays which monitor the kinetics of amyloid fibril growth were extensively performed in this thesis. The kinetics of amyloid fibril formation from bovine insulin and ovalbumin were performed in FluoStar Optima

and Omega plate readers. These plate readers are filter-based fluorescence plate readers. The filters used were a 450 nm excitation and 485 nm emission with a band width of ± 10 nm.

2.2.1.2 Insulin Assays

Bovine insulin was obtained from Sigma-Aldrich (I5500, lot # 0001434060). The zinc content was approximately 0.5 % (w/w). The samples in this study were dissolved in 25 mM HCl (pH 1.6) immediately prior to the experiment. All solvents and solutions were filtered through a 0.22 μm filter (Millipore). Concentrations were checked via UV-Vis absorption spectroscopy. ThT was added to each solution to a final concentration of 20 μM. Experiments were conducted using Corning NBS 96-well plates (Corning 3641). These plates are coated with a proprietary PEO-like hydrophilic coating which minimizes the interaction of protein with the plates. It was found that using these plates significantly minimized the variability in the kinetics compared to standard polystyrene microwell plates. Each well of the plate was filled with 100 μL of solution. Experiments for each protein concentration were replicated across 2-3 whole plates. The plates were sealed with a plastic adhesive and then incubated at 60° C. Fluorescence readings were taken from the bottom optic. Readings were taken every 10 minutes for all kinetic experiments. In the case of experiments with agitation, the plate was shaken at 600 rpm for two minutes. It then remained quiescent for the remaining eight minutes of the cycle. The plate was not shaken for the entire duration of the cycle so as to minimize the effects of any temperature gradients within the plate reader (for a discussion on temperature inhomogeneities within plate readers see Appendix A). The final number of individual experiments for a given protein concentration ranged from $\sim$ 140-200.

The experiments involving adding NaCl at regular time points during the course of fibril formation (Chapter 3.2.5) were performed by preparing a microwell plate of 1.5 mg/ml insulin samples containing 20 μM of ThT. 6 μL of a 5M stock NaCl solution was added to three wells at the start of the experiment. The plate was placed into the plate reader at 60°C and ThT fluorescence was monitored. Subsequently, 6 μL of the same concentrated NaCl solution was added to another three wells every 30 minutes, for a total of 240 minutes. A final reading was taken at 400 minutes. The final concentration of NaCl in the wells was 0.3M.

Concentrated NaCl was added in small volumes to minimize any change in volume which could affect the kinetics of fibril formation.

2.2.1.3 Ovalbumin Assays

Albumin from chicken egg white was obtained from Sigma-Aldrich in two grades of differing purities: Grade 2 (A5253) and Grade 5 (A5503). Ovalbumin was incubated in the presence of 10mM dithiothreitol (DTT) and 10 mM Ammonium Acetate. ThT was added to a final concentration of 55 μM. Samples were initially incubated at 37°C for 2 hours. This is following the protocol outlined by [144]. This step is performed in order to reduce intra-protein disulphide bonds which was found to increase the rate of formation of amyloid-like fibrils. After this pre-incubation, samples were added to microwell plates at a volume of 100 μL. Plates were shaken at 600 rpm (double orbital) for 1 minute at a temperature of 60° C. Total time between cycles was 2 minutes. The experiments measuring the kinetics of fibril formation were performed in a variety of plates: Corning NBS 96 and 384 well plates, as well as Iwaki 96 and 384 polystyrene microplates. Microwell plates were sealed with a plastic adhesive. Temperature dependence studies were performed on a Varian Cary Eclipse Fluorimeter in quartz cuvettes. Three cuvettes were monitored for each temperature studied.

2.2.2 UV-Vis Spectroscopy

Insulin concentrations were found using a Cary UV-Vis spectrophotometer using Varian software. Samples were placed in a quartz cuvette and a spectral scan was taken for wavelengths ranging from 200 to 500 nm at a scan rate of 30 nm/sec. The spectrum was blank corrected using a matching quartz cuvette and buffer. The extinction coefficient of bovine insulin is 5734 $M^{-1}cm^{-1}$. For the experiments determining the amount of insulin left in solution after fibrils have formed, fibril samples were centrifuged in eppendorf tubes at 18,000 rpm for 30 minutes. The supernatant was carefully pipetted away without disturbing the pellet. The supernatant was then scanned as described above.

2.2.3 Bradford Assay: Another Means of Determining Protein Concentration

Bradford Assays are a means of determining the concentration of a protein in solution when performing UV-vis absorbance analysis is difficult or not possible. Unlike UV-vis spectroscopy where an absolute value for the concentration is found via the intrinsic absorption of the protein, the Bradford assay determines concentration relative to a known standard concentration curve. The key to the Bradford Assay is a spectral shift in absorbance of Coomassie Blue dye when bound to protein. Coomassie Blue, at very acidic conditions, exists as a stable, doubly-protonated form which shows characteristic absorbance $\sim$ 470 nm (appears red/brown in color). However, upon binding protein, Coomassie blue becomes unprotonated and shifts to an absorbance of $\sim$595 nm (becomes blue). The Beer-Lambert law is then applicable in determining the concentration of protein based on the amount of absorbance at 595 nm. Bradford Assays are most reliable when comparing the absorbance shift of Coomassie Blue with a known standard. Bovine serum albumin (BSA) is a common protein used as a Bradford Assay standard. In practice, one makes a serial dilution of BSA and measures the absorbance at 595 nm in the presence of Coomassie Blue. For sufficiently low concentrations, one will obtain a linear trend in the absorbance. These values can be fit by linear regression to form a standard curve. One can then measure the Coomassie Blue absorbance for an unknown concentration of protein and utilize the standard BSA curve to determine the concentration.

This technique was used when determining the amount of unaggregated ovalbumin present after incubating 24 hours in conditions promoting the formation of amyloid fibrils. This was done because the reducing agent, DTT, has a similar absorbance band to proteins. Bradford Assays were performed to avoid this complication. The ovalbumin amyloid fibril samples were spun in centrifuge tubes at 3200 rpm for 30 minutes and 4°C. The supernatant was then removed. A standard curve of BSA was made by making a serial dilution from 1.5 mg/ml down to 0.125 mg/ml. Bradford reagent was added to the BSA dilution and then three replicates of each concentration was aliquotted to a 96 well plate. The absorbance at 595 nm was determined by a BMG FluoStar plate reader. The absorbance values for each replicate were averaged and a standard curve was generated by fitting the data by linear regression. Bradford reagent was

then added to the supernatant and added to a plate reader in triplicate. Again, absorbance readings were taken at 595 nm which were averaged. The values were then compared to the standard curve to determine the concentration.

2.2.4 Transmission Electron Microscopy

Transmission electron microscopy (TEM) is a valuable tool for imaging objects on the nanometer scale. This is because the de Broglie wavelength of an electron in a typical TEM instrument (accelerating potential $\sim$ 100 kV) is on the order of 10^{-12}m. An image is acquired through the interaction of an electron beam with the sample. In standard 'bright field' mode, an image which absorbs more electrons will stand out as a darker object relative to one that absorbs electrons less. In many cases, particularly for biological samples, a negative stain such as uranyl acetate is used to enhance contrast.

2.2.4.1 TEM of Amyloid Fibrils

Insulin and ovalbumin amyloid fibrils were made by incubating samples in the conditions outlined above. Fibril samples were then diluted to either 1:5, 1:10, or 1:20, depending on the concentration of the starting sample. This dilution is carried out because concentrated samples are too dense to image. Aliquots of 3 μL fibril samples were placed on 400 mesh copper grids (Taab, Aldermaston, U.K.) coated with formvar and carbon and allowed to left for 5 minutes. The grids were washed with 10 μL MilliQ water followed by application of 4 μL of 2% uranyl acetate (w/v). The uranyl acetate was left on the sample grid for 30 seconds. Excess water and stain were removed with filter paper and the grids were left to air-dry for 20 minutes. Imaging was performed using a Philips (FEI) CM120 Biotwin transmission electron microscope with an operating voltage of -80kV.

Length distributions were obtained using ImageJ image analysis software. Using a sufficient dilution (1:20), one can obtain samples which contain fibrils that do not excessively overlap and are easily distinguishable as individuals. Fibrils were then traced and individual lengths obtained. Fibrils used in the analysis were measured over several images in various regions of the grid and using different samples.

2.2.5 Microrheology: Bead Tracking

2.2.5.1 Brownian Motion

Brownian motion, named after Robert Brown who noted the chaotic, non-directed motion of pollen in water in 1827, is the stochastic movement of particles suspended in solution. The correct interpretation of Brownian motion was given by Einstein in 1905-that the random motion is due to thermally driven collisions between the particle and the constituent solution. Einstein made the insight that the translational displacement of a Brownian particle can be described by a diffusion equation,

$$\frac{\partial \rho}{\partial t} = D\frac{\partial^2 \rho}{\partial x^2} \tag{2.1}$$

where, $\rho(x,t)$ is the density of Brownian particle at position x and time t, and D is the diffusion constant. The solution to this equation is a Gaussian function describing the spatial and temporal evolution of the density of Brownian particles,

$$\rho(x,t) = \frac{1}{(4\pi Dt)^{1/2}} e^{-\frac{x^2}{4Dt}} \tag{2.2}$$

Given this probability distribution function, the mean displacement may be calculated. It is readily evident from the nature of the Gaussian function that $< x >= 0$. The second moment, the mean square displacement, is non-zero $< x^2 >= 2Dt$. This result can be extended to any dimension, d, such that

$$< \Delta r^2 >= 2dD\Delta t \tag{2.3}$$

This result shows that the mean square fluctuations of a particle depends *linearly* on time. This result is a hallmark of isotropic random-walk processes. This linear temporal dependence on the mean square displacement is true for the case of diffusion in a viscous fluid. The existence of an elastic component to the fluid in which the particles are diffusing results in sub-linear time dependence [152, 153], i.e., $< \Delta r^2 >= 2dD\Delta t^{\alpha}$, where $\alpha < 1$, over long time scales. It should be noted that this sub-linearity of α is theoretically transient and in the long time limit, $\alpha \to 1$.

Einstein made a second and important observation: that the forces that drive the fluctuation of Brownian particles are the same dissipative forces that exert a viscous drag on the particles. This relation, known generally as the fluctuation-

dissipation theorem, relates the diffusion constant, D, and the mobility of particles in solution. He showed that for a spherical particle the diffusion constant can be written as,

$$D = \frac{kT}{6\pi\eta a} \tag{2.4}$$

where k is Boltzmann's constant, T is the temperature, η is the viscosity of the solution, and a is the radius of a diffusing particle. Therefore, by measuring the diffusion constant one can determine the viscosity of a substance, temperature of the system or the size (and shape) of diffusing particles.

2.2.5.2 Video Particle Tracking

From the insights gained above, video particle tracking allows one to monitor the diffusion of particles within a potentially complex fluid system and thereby determine important material properties about that system. A typical experiment consists of recording the motion of tracer particles (typically spherical beads with radii of 1 μm or less) within a fluid of interest using a microscope that has been connected to a CCD camera where images may be acquired at a desired frame rate.

These raw images can then be analyzed using a number of computer programs. In this thesis, particle tracking was achieved using the very popular algorithm developed by Crocker and Grier for IDL [154]. The algorithm identifies candidate particles, locates their position, and then, frame-by-frame, links the particles into a trajectory. The information obtained provides the x- and y- position of each particle within the field of view for each frame.

This information can then be used to calculate the mean squared displacement (MSD) of each particle. For simplicity consider the one-dimensional case. Call x_i the position at t_i and the interval between 2 successive frames Δt. Any subsequent position $x_{i+1} = x(t_i + \Delta t)$ and therefore any displacement $\Delta x_i(\Delta t) = x_{i+1} - x_i$. Thus, the MSD for that given time interval is obtained by averaging all steps, S_n, corresponding to that interval,

$$< (\Delta x(\Delta t))^2 >= \sum_{i=1}^{n} x_i^2(\Delta t) \tag{2.5}$$

This is the MSD for only one frame interval, e.g. 1/100 s for a frame rate

of 100 fps. One wants to plot the MSD for many time intervals. Therefore, one must then calculate $< (\Delta x_i(n\Delta t))^2 >$ for n time intervals. Of course, for longer time intervals, the number of steps decreases, resulting in increasing statistical error. Therefore, it is beneficial to calculate a large ensemble average of diffusing particles to minimize the statistical errors at longer times. Additionally, taking large ensemble averages can improve the error in the MSD due to polydispersity in the size or shape of tracer particles.

2.2.5.3 Particle Tracking: Amyloid Fibrils

Insulin amyloid fibril samples, at several concentrations, were formed in NBS 96-well microwell plates at 60°C without shaking for 24 hours. The plates were then removed and allowed to equilibrate to room temperature. A solution of latex beads, 1 μm in diameter, were diluted to 1:5 in the buffer used in the formation of the insulin amyloid fibrils. Roughly 1 μL of the latex bead solution was added to a microwell containing the amyloid fibrils. The beads were initially gently mixed with the tip of the pipette. The fibril and bead samples were removed by pipette and $\sim$ 100 μL was placed onto a glass microscope slide. A glass cover slip with petroleum jelly applied to its four edges was carefully placed onto the glass slide containing the sample. The glass cover slide was gently pressed onto the lower glass slide making an airtight seal. The sample was then imaged using a Nikon Ti-U inverted microscope with a Nikon 40x ELWD objective. The images were captured using a Mikrotron EoSens CCD camera at 100 fps over a field of view of 500x500 pixels. The movies were saved as individual .bmp files. For each sample, several movies in different regions of the sample were acquired, with each movie containing at least 50 beads within the field of view. These files were then converted into tiff stacks using ImageJ software. Once converted to tiff stacks, background subtraction, image inversion (so that there are white beads on a black background), and contrast adjustments were performed. This was done because the IDL tracking routines work best for white, bright 'Gaussian' blobs on a flat black background. Particle identification, tracking, drift removal, and the calculations of the MSD were performed using IDL routines of [154].

2.2.5.4 Particle Tracking: Ocr

To determine the physical characteristics of the aggregated ocr variants, microrheological bead tracking experiments were performed. Experiments using bulk rheological techniques was not a viable option because of the prohibitively small quantities of sample available. The protein variants were stored in a -20C freezer in 50% (by weight) glycerol. The glycerol was removed and protein exchanged into a 20mM Tris-HCl, pH 8 buffer using NAP-5 desalting columns (GE=Healthcare). After buffer exchange, sample concentrations were determined using UV-Vis absorbance (extinction coefficient of 31860$M^{-1}cm^{-1}$). Due to varying degrees of dilution during buffer exchange, some samples required concentrating. This was done using a centrifugation concentrator (Vivaspin-2). A final concentration for all protein samples was set to be 36μM. Latex, sulphate-free 2 μm beads were added to the samples to a final volume fraction $\sim$ 0.01%. Samples were then placed in microdialysis buttons (Hampton Research) and sealed with an equilibrated dialysis membrane with a MWCO of 10 kDa. Buttons were then placed into a range of ammonium sulphate concentrations. A stock ammonium sulphate solution of 3.5M was made by dissolving ammonium sulphate (Sigma-Aldrich) into 20 mM Tris-HCl pH 8. Subsequent concentrations were made by dilution using 20 mM Tris-HCl pH 8. The samples were allowed to dialyze between 4-6 hours at 4°C in a cold room. Prior to imaging, the samples were removed and allowed to equilibrate to the temperature of the room where the imaging was taking place (30 minutes). The samples were then directly imaged through the button by placing the button onto a glass cover slip. Images were acquired using a Nikon Ti-U inverted microscope with a Nikon 40x ELWD objective. The images were captured using a Mikrotron EoSens CCD camera at 100 fps over a field of view of 500 x 500 pixels. Analysis of the subsequent images to obtain MSD data follows as described above.

2.2.6 Circular Dichroism Spectroscopy

Circular dichroism spectroscopy utilizes the effect that optically active chiral molecules will differentially absorb left- or right-circularly polarized light (LCP and RCP) [155]. Linearly polarized light can be thought of as a superposition between LCP and RCP light with equal amplitude and phase. This can be

visualized by viewing the projection of the two components, orthogonal to the direction of propagation (see Fig. 2.5A). However, when light passes through a material that differentially absorbs the two components, the result is a difference in the amplitudes of these two components. When looking at the projection again, one finds instead of a line, an ellipse is formed (Fig. 2.5B). This ellipticity is known as circular dichroism (CD). A useful measure of CD is the degree of ellipticity which is defined as the tangent of the ratio of the major and minor axes of the ellipse (θ in Fig. 2.5C). It should be noted that the CD signal in most samples is extremely small (on the order of milli-degrees), and the ellipticities shown in Fig. 2.5B are only for illustrative purposes.

Circular dichroism is particularly useful for discerning secondary and tertiary structure of proteins. For a CD signal to be observed, radiation must be absorbed. Proteins contain multiple chromophores of interest including the peptide bond (240 nm) and the aromatic residues (260-320 nm) in addition to co-factors such as haem groups that show characteristic absorption bands. Moreover, when many chromophores of a particular type are spatially arranged in close proximity to one another they may act as a single absorber which can result in a distinct spectrum. CD is therefore particularly suited in providing complementary structural information over a range of spectral regions.

In particular, absorption bands in the far-UV(180-240 nm), provide information about the secondary structure of a protein. This is primarily due to the absorbance and spatial arrangement of the peptide bonds within a protein. For instance, α-helices, β-sheets, β-turns, or random coiled structures have distinct CD spectra (see Fig. 2.6). Indeed, many computer algorithms have been developed to analyze CD spectra to determine the percentage of α-helix or β-sheet within a protein (see for instance [156]). Information about the tertiary structure of a protein can be obtained by looking at the CD spectra in the near-UV (260-320 nm). In this case the primary chromophores are the aromatic residues. A CD spectrum in the near-UV is determined by the number and identity of the aromatic residues within a protein, as well as their spatial arrangement, proximity, and mobility [155]. Due to this, the solution or environmental conditions may have a marked effect on the CD spectrum. The CD spectra obtained in the near-UV often can not be ascribed to a particular structure like in the far-UV case. However, near-UV spectra are particularly useful in discerning changes in

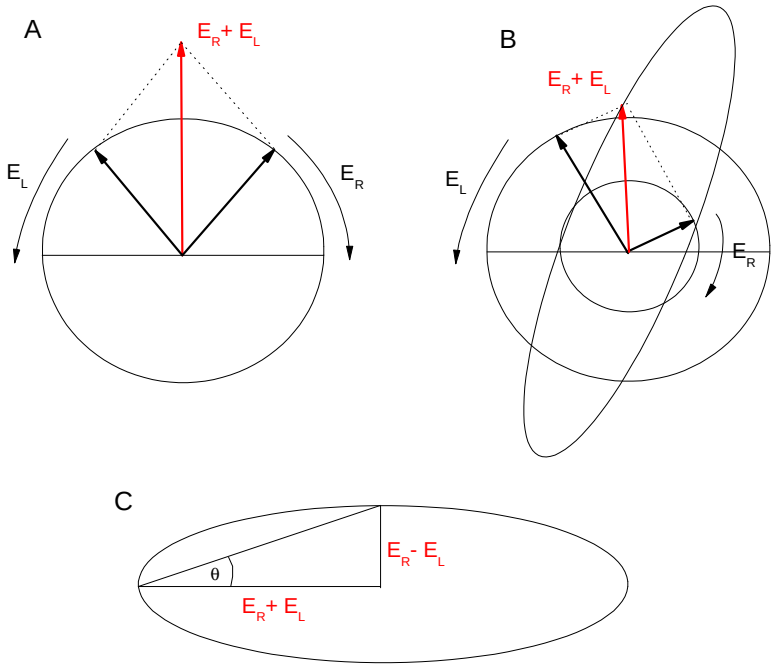

Figure 2.5: **Origin of circular dichroism.** **A** Projection of the left- and right- circularly polarized light perpendicular to the direction of propagation. Linearly polarized light can be thought of as a superposition of left- and right-circularly polarized light that have the same amplitude and phase. **B** However, if light passes through an optically active medium which differentially absorbs left- and right- circularly polarized light (e.g. chiral molecule), the component which is absorbed more will experience a reduction in amplitude; projecting these components orthogonally to the propagation direction results in tracing out an ellipse. This ellipticity is known as circular dichroism (CD). When there is circular dichroism, there always exists a change in the phase between the left- and right-circularly polarized components. This results in a change in the polarization plane (the elliptical axes) creating a 'tipping' of the ellipse by some angle. **C** Circular dichroism experiments often report the amount of circular dichroism in a sample by measuring the ellipticity, θ, which is the tangent of the ratio of the minor and major axes of the ellipse (usually measured in milli-degrees).

tertiary structure of a protein, for instance, between a wild-type and mutant variant. CD spectroscopy is a very useful tool in gaining general information about the secondary and tertiary structures of peptides and proteins.

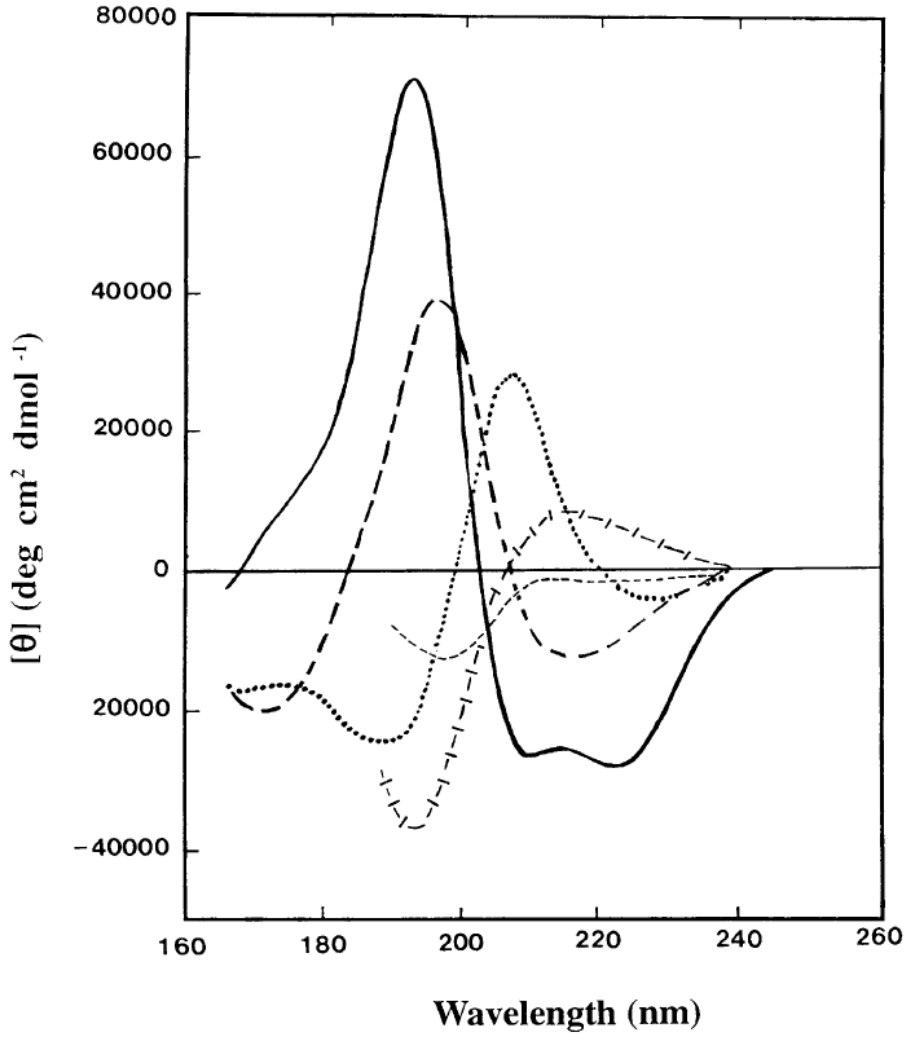

Figure 2.6: **Standard far-UV CD spectra for common protein secondary structure motifs.** One can determine the presence and the relative amount of a particular type of secondary structure due to these motifs having distinct CD 'finger prints'. Some examples are α-helix (solid line), anti-parallel β-sheet (long dashed line), type I β-turn (dotted line), extended 3_{10}-helix (cross dashed line), irregular structure (short dashed line). Taken from [5]

2.2.6.1 CD of Insulin Amyloid Fibrils

A range of insulin solutions between 0.25 and 4 mg/ml were incubated at 60°C without shaking for 24 hours in NBS 96 microwell plates. CD spectroscopy of the resulting fibrils was performed on a Jasco J-810 spectropolarimeter. The

spectropolarimeter was purged with N_2 gas for 24 hours prior to the experiment. 75 μL of fibril samples were placed in a quartz cuvette with a path length of 0.1 mm. Spectra were obtained in the far-UV (190-230 nm) by averaging over 5 accumulations. A bandwidth of 1nm and sensitivity of 100mdeg was used in all experiments. The J-810 gives the ellipticity in millidegree units. The data was converted to molar ellipticity (deg cm^2 $dmol^{-1}$) which takes into account differences in concentration between samples.

2.2.7 Neutron Scattering

There are various techniques which can be employed to elucidate the molecular organization of biopolymers. X-ray diffraction spectroscopy is a powerful tool in this regard, capable of providing spatial information for every atom within a molecule. However, this technique suffers from several drawbacks. Firstly, such high molecular resolution is achieved through single crystal x-ray diffraction. It is a very non-trivial experimental hurdle to produce high quality single crystals of many biological molecules. Secondly, the use of ionizing radiation can cause irreparable damage and eventual destruction of a sample. Indeed, with the modern high luminosity synchrotron X-ray sources available at many international laboratories this consideration must be accounted for.

Neutron scattering techniques avoid these issues altogether. Neutron scattering experiments can be performed in a myriad of conditions (e.g. in complex solutions and solvents) with effectively little or no damage to the sample. The origin of these differences lies in the simple fact of how X-ray and neutron radiation interacts with a sample: x-rays are scattered by electrons, neutrons are scattered by nuclei. The strength of the interaction between X-rays and a given atom, described by the 'scattering length', is a function of the atomic number. This is not the case for the neutron-nucleon interaction. Indeed, hydrogen, which is usually not resolvable in X-ray diffraction experiments, has an (incoherent) scattering length almost an order of magnitude larger than any other nuclei. Even more importantly, the scattering lengths of hydrogen and deuterium are markedly different. Because of this fact, it is typically advantageous to perform neutron scattering experiments in D_2O or mixtures of H_2O and D_2O to minimize the large incoherent scattering background present in solutions containing H_2O as a solvent.

2.2.7.1 Basics of Neutron Scattering

In this section a very brief overview of scattering, with an emphasis on small angle neutron scattering, will be provided in order to elucidate some of the results presented later. The task at hand is to understand how the information obtained from scattering a neutron from a sample (i.e. changes in wavevector and energy) relates to structural or dynamical information.

While neutrons are often thought in terms of their particle nature, they of course possess a wavelength according to the de Broglie formula $\lambda = h/mv$ where h is Planck's constant, m is the mass of the particle and v its speed. It should be noted, thermal neutrons, that is neutrons with energy kT, where T is approximately room temperature, have wavelengths on the order of an angstrom. This is ideal for probing the structures relevant to molecular characteristics of biopolymers. Scattering events can be cast in terms of neutron wavevectors, $\mathbf{k}$. The wavevector $\mathbf{k}$ has magnitude $2\pi/\lambda$ and points in the direction of motion. In an elastic scattering event, the wavelength λ is unchanged only the direction of $\mathbf{k}$ is altered. This change can be parameterized by the scattering vector $\mathbf{Q} = \mathbf{k_f} - \mathbf{k_i}$ where $\mathbf{k_f}$ and $\mathbf{k_i}$ are the final and initial scattering vectors, respectively (Fig. 2.7). The scattering vector $\mathbf{Q}$ has magnitude

$$Q = \frac{4\pi}{\lambda} sin(\theta/2) \tag{2.6}$$

Therefore, Q encodes the spatial information, in real space, of a scattering sample.

Generally, consider a sample consisting of N nuclei, each of which has some scattering length b_k. Recall, scattering length describes how strongly a particular form of radiation will interact with some nuclei. More concretely, the scattering length, b can be interpreted as the amplitude of a scattered wave by some nuclei. With this in mind, one can write down the form of the scattered waves from these N nuclei as

$$\mathbf{\Psi}(\mathbf{r}) = \sum_{k}^{N} \frac{b_k}{r_k} \exp(i\mathbf{Q} \cdot \mathbf{r_k}) \tag{2.7}$$

where $\mathbf{r_k}$ is the distance of the kth atom from an observer and $\mathbf{Q}$ is the scattering vector. Since $\mathbf{r_k} \gg \mathbf{r_j} - \mathbf{r_k}$, one can write $\mathbf{r_k}$ simply as r.

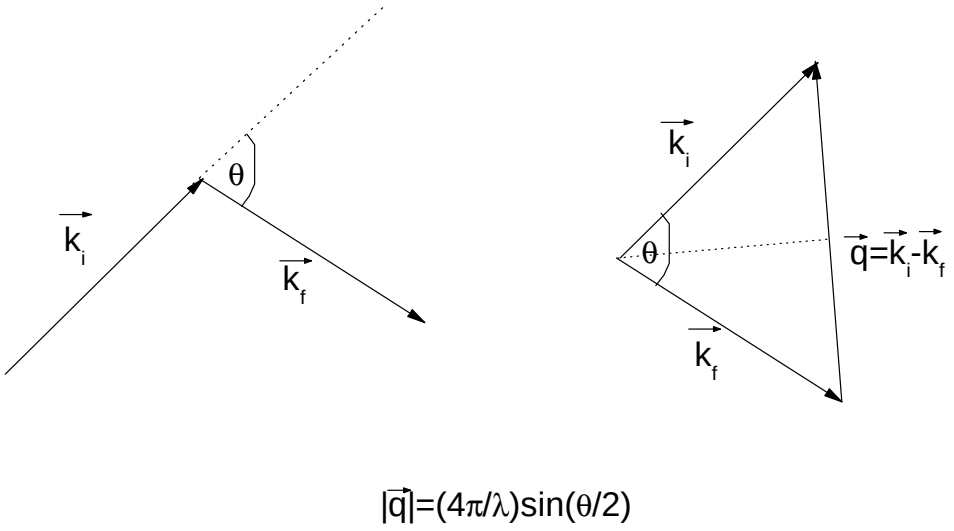

Figure 2.7: **Graphical origin of the Q-vector.** An incoming particle with wave vector $\mathbf{k_i}$ (magnitude of $2\pi/\lambda$) scatters by some angle θ and the resulting wave vector is $\mathbf{k_f}$. If the scattering event is elastic, there is no energy exchange between the particle that's been scattered and the scattering center (e.g. the nucleus). In that case, the wavelength is unchanged and the magnitude of $\mathbf{k_i}$ and $\mathbf{k_f}$ are the same; only the direction changes. This change in direction is quantified by the scattering vector $\mathbf{Q}$ whose magnitude is $Q = \frac{4\pi}{\lambda} sin(\theta/2)$.

One can consider a small solid angle dΩ in which the number of neutrons, per second, are measured. This number will depend on the orientation of dΩ, the energy, E of the neutrons, and the initial intensity of the incoming neutron beam, I_0. This number is known as the partial differential cross section and is given by,

$$I_0 \frac{\partial^2 \sigma}{\partial \Omega \partial E} d\Omega dE \tag{2.8}$$

Integrating over the final energies gives the differential cross section $\partial\sigma/\partial\Omega$ which gives the total number of neutrons scattered into dΩ per second. In other words, the number of neutrons per solid angle per time (i.e the intensity) is given by

$$I = I_0 \frac{\partial \sigma}{\partial \Omega} \tag{2.9}$$

Of course, the intensity is equivalently given by $\langle \Psi \Psi^* \rangle$, therefore,

$$\frac{\partial \sigma}{\partial \Omega} = \langle \sum_k^N b_k \exp(i\mathbf{Q} \cdot \mathbf{r_k}) \sum_{\mathbf{j}}^{\mathbf{N}} \mathbf{b_j} \ \exp(-\mathbf{iQ} \cdot \mathbf{r_j}) \rangle \tag{2.10}$$

$$= \sum_{j,k}^N \langle b_j b_k \ \exp(i\mathbf{Q} \cdot (\mathbf{r_k} - \mathbf{r_j})) \rangle \tag{2.11}$$

This sum can be partitioned into two parts: where $j = k$ and where $j \neq k$,

$$\frac{\partial \sigma}{\partial \Omega} = \sum_{k=1}^N b_k^2 + \sum_{j \neq k}^N \langle b_j b_k \ \exp(i\mathbf{Q} \cdot (\mathbf{r_k} - \mathbf{r_j})) \rangle \tag{2.12}$$

which results in

$$\frac{\partial \sigma}{\partial \Omega} = N\langle b^2 \rangle + \langle b \rangle^2 \sum_{j \neq k}^N \ \exp(i\mathbf{Q} \cdot (\mathbf{r_k} - \mathbf{r_j})) \rangle \tag{2.13}$$

The term $\langle b \rangle^2$ results due to the fact b_j and b_k are uncorrelated with $\mathbf{r_j}$ and $\mathbf{r_k}$. We can define the fluctuations in b as $\overline{\Delta b^2} = \langle b^2 \rangle - \langle b \rangle^2$. The differential cross-section then becomes,

$$\frac{\partial \sigma}{\partial \Omega} = N\overline{\Delta b^2} + \langle b \rangle^2 \sum_{j,k}^N \langle \exp(i\mathbf{Q} \cdot (\mathbf{r_k} - \mathbf{r_j})) \rangle \tag{2.14}$$

The differential cross-section has been partitioned into two parts: one which only depends on the fluctuations in the scattering length, b, and the second depends on the relative positions of scattering nuclei. Incoherence arises due to scattering centers with different scattering lengths. This could arise due to the presence of isotopic impurities in a sample, for instance. Indeed, the second term gives an angular dependence to the scattering intensity and encodes the structural information of sample. We can therefore define the structure factor $S(Q)$,

$$S(Q) = \frac{1}{N} \langle \sum_{j,k}^N \exp(i\mathbf{Q} \cdot (\mathbf{r_k} - \mathbf{r_j})) \rangle \tag{2.15}$$

We can define the coherent scattering cross section as $\sigma_{coh} = 4\pi\langle b\rangle^2$ and the incoherent scattering cross section, $\sigma_{inc} = 4\pi\overline{\Delta b^2}$. For the case of hydrogen, one finds experimentally that σ_{coh}=1x10^{-24}cm^2 and σ_{inc}=80x10^{-24}cm^2. The coherent scattering cross-section for hydrogen is nearly an order of magnitude larger than any other nuclei. Indeed, the coherent cross section of hydrogen is larger than all other nuclei. This important fact is the foundation for hydrogen/deuterium contrast. The experimentalist, for instance, may exploit the difference in scattering between a D_2O solvent and a hydrogenated sample in order to reduce the large incoherent scattering background one would have if using H_2O as a solvent. Another example would be to selectively deuterate a region of polymer, thus providing scattering contrast between that region and the rest of the polymer.

In practice, one measures the intensity of neutrons received by a detector as a function of the scattering vector, Q. At values of low Q one can write an approximate form of $I(Q)$,

$$I(Q) = I(0)\ \exp(-(qR_g)^2/3) \tag{2.16}$$

where R_g is the radius of gyration. Writing $I(Q)$ in this form is known as the Guinier approximation. By plotting $lnI(Q)$ vs. Q^2 one can obtain R_g by simply calculating the slope of the linear plot. Clearly, Guinier analysis can be extremely useful for determining the size of structures within a sample-particularly when coupled with time-resolved scattering. One must take caution, however, when calculating R_g in the Guinier regime for it is only valid for values of $qR_g \ll 1$.

At larger Q-values (i.e. smaller length scales), one can consider the scattering of neutrons from a 'fractal' surface. In practice, fractal objects that can be observed via neutron scattering are gaussian chains with excluded volume, diffusion limited aggregates, or gel networks. It can be shown that the intensity scales with fractal dimension as $I(Q) \approx Q^{-D}$, where D is the fractal dimension of the surface. This analysis can be useful, for instance, when studying the formation and structure of gels; gels may be characterized by the fractal dimension of the clusters that constitute them. As ever, one must be careful using this type of analysis. When plotting $lnI(Q)$ vs. lnQ, the linear regime must extend over at least one decade of Q to obtain any reliable information on the system in question.

2.2.7.2 ISIS Neutron Source and the Sans2d Detector

Neutron scattering experiments performed in this thesis was carried out at the ISIS facility which is located at the Rutherford-Appleton Laboratory in Harwell, England. ISIS produces neutrons by first accelerating H^- ions to 70 MeV by a linear accelerator. These ions are then transported to a synchrotron, where, upon entry, they are stripped of their electrons. The resulting protons are accelerated to an energy of 800 MeV. These protons are extracted from the synchrotron and delivered to two different target stations which produce neutron beams through the spallation process. Experiments in this thesis were performed at Target Station 2, which is designed to provide long-wavelength and low energy neutrons. The neutrons produced are ideal for studying topics in soft matter and biological physics.

Experiments were performed on the Sans2d detector. The Sans2d instrument measures over a Q-range of 0.002-3 $Å^{-1}$. The detector consists of two ORDELA (Oak Ridge, Tennessee, USA) two-dimensional area detectors filled with ^{3}He-CF_4 gas. The active area on each detector is 96.5 cm x 96.5 cm with 5 mm resolution. The detectors can be moved, in vacuum, so that the detector distance can vary between 2 to 12m. To further increase the Q-range, for a given detector distance, the detectors can be laterally offset up to 1200 mm.

2.2.7.3 Neutron Scattering of Insulin Amyloid Fibrils

Insulin samples were prepared using zinc-free bovine insulin purchased from Sigma-Aldrich (I5500). Insulin samples were freshly prepared prior to every experiment. Insulin was dissolved in fresh D_2O and pH adjusted with HCl to a final value of pH 1.6. The samples were then filtered using a 0.22μm polyether sulfone membrane filter. To ensure no pre-formed aggregates were remaining in the sample, DLS was performed. It was confirmed prior to the experiments no pre-formed aggregates were present. Both detectors were positioned at a distance of L_1=L_2=4m. Samples were placed in quartz 'banjo' cells with a thickness of 2mm and diameter of 20mm. The Q-range covered in these experiments was 0.005-0.77 $Å^{-1}$. Calculations involving molecular weight were found using the equation $M_w = I(0)N_A/c(\Delta\rho)^2\nu^2$, where N_A is Avogadro's number, c is protein concentration (gm/cm^{-3}), $\Delta\rho$ is the contrast difference between solvent and protein (3.4x10^{10} cm^{-2}), and ν is the partial volume of the protein (0.73 cm^{3}g^{-1}).

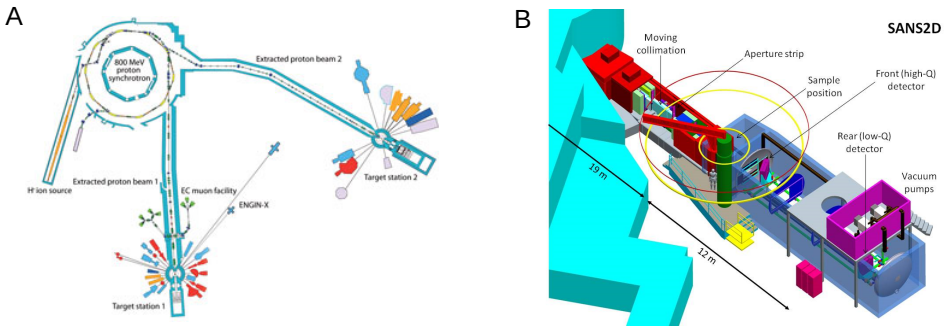

Figure 2.8: **Schematic of ISIS neutron source and the Sans2D detector.** The ISIS neutron source creates high energy protons (800 MeV) via a linear and synchrotron accelerator. Neutron beams are created by colliding these protons into tungsten plates (spallation). There are two target stations at ISIS; Target Station 2 is optimized to deliver longer wavelength and lower energy neutrons ideal for soft matter and biological experiments. The Sans2d detector is a time-of-flight, small angle neutron scattering instrument. It contains two detectors allowing for a typical Q-range coverage of 0.002-3A^{-1}.

The collected data was fitted using SasView software. The data are fit with two different models. The first is model which assumes the scatterers are of a cylindrical geometry. The model provides the form factor for a circular cylinder with uniform scattering length density. The 2-D scattering intensity for oriented cylinders is given by,

$$I(Q,\alpha) = \frac{A}{V} f^2(Q) + b \tag{2.17}$$

where,

$$f = 2\Delta\rho V sin\left(QL\frac{cos\alpha}{2}\right) \Big/ \left(QL\frac{cos\alpha}{2}\right) \frac{J_1 Qrsin\alpha}{Qrsin\alpha}. \tag{2.18}$$

The angle α is the angle between the axis of the cylinder and the Q-vector, V is the volume of the cylinder, L is the length and r the radius of the cylinder, $\Delta\rho$ is the scattering length density difference between the scatterer and solvent (contrast), b is the background, A is a scale factor and J_1 is a first-order Bessel function. For fitting, the axis of the cylinder is defined by two angles, θ and ϕ (see Fig. 2.9 for their definition). The returned fit parameters are A, r, L, $\Delta\rho$, b, θ, and ϕ.

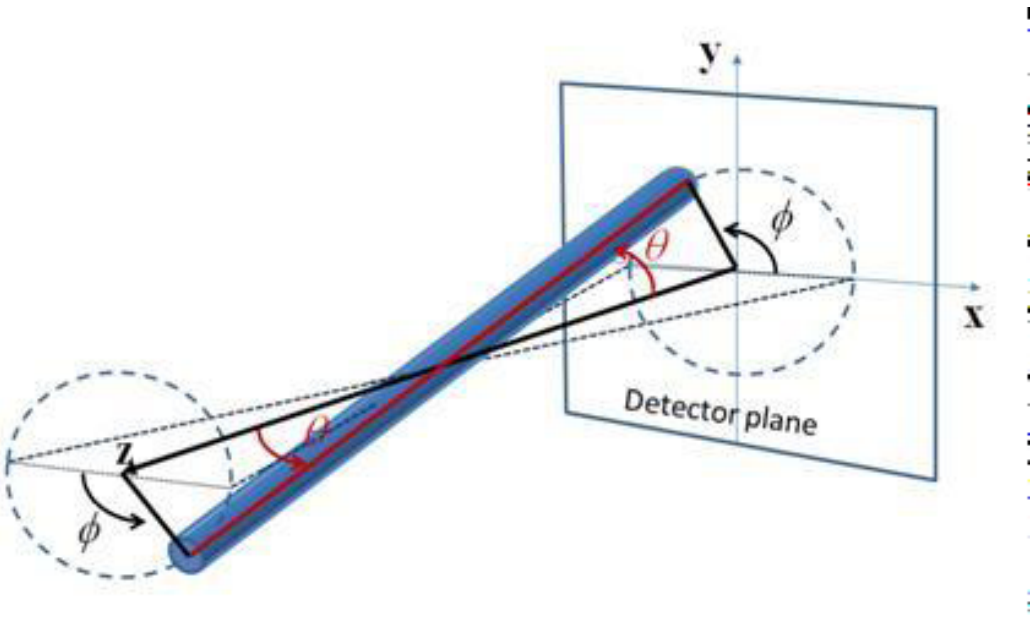

Figure 2.9: **Definition of θ and ϕ used in the fitting of cylinder model to data.** (Image taken from SasView Manual [6].)

The second model used to fit the data is simply a linear combination of two cylinder models.

The scattering intensity $I(Q)$ is related to the real-space length distribution of scatterers, $P(r)$, by a Fourier transform,

$$I(Q) = 4\pi \int_0^{D_{max}} P(r) \frac{sin(Qr)}{Qr} \mathrm{d}r \tag{2.19}$$

where D_{max} is the maximum dimension of a scattering unit. In practice, a finite Q-range with discrete Q-values are explored in an experiment, therefore the integral above can not be solved exactly. Algorithms have been developed to indirectly measure $P(r)$ from $I(Q)$. SASview calculates $P(r)$ by using the method of [157]. In essence, the distribution $P(r)$ is related to a discrete Fourier sine series. The components of this series are calculated by performing a least-squares fit to the measured $I(Q)$ signal. The parameters that are free to be chosen are the number of terms in the Fourier sine expansion, a regularization term which smooths the resulting $P(r)$ output, and the maximum distance D_{max}. It should be noted that the values of $I(0)$ and R_g as outputs since $I(0)$ is simply the area under the obtained distribution and R_g is simply the second moment of

$P(r)$.

2.2.8 Brief Overview on Stochastic Simulations of Amyloid Fibril Formation

In this thesis, in order to interpret the kinetic experimental results we used coarse grained stochastic simulations that utilize the Gillespie algorithm [158]. The simulations were coded and executed by K. Eden. Our simulation model consists of the following set of chemical reactions:

$$f_i + m \xrightarrow{2k_+} f_{i+1} \tag{2.20}$$

$$f_i \xrightarrow{k_f} f_j + f_{i-j} \tag{2.21}$$

Here, f_i denotes a fibril of length i and m denotes a protein monomer. We assume that the shortest possible fibril length is 2 monomer units (i.e. the "nucleus size" is 2). Reaction (1) represents elongation by monomer addition, with rate constant k_+ (the factor of 2 arises from the fact that monomers can be added at either end of a fibril). Reaction (2) represents fragmentation of a fibril into two shorter fibrils: k_f is the fragmentation rate, which we assume to be independent of the position along the fibril at which fragmentation happens. Note that this reaction also incorporates monomer loss from the fibril ends (here $f_1 \equiv m$). Nucleation is not explicitly modeled in our system because the kinetics of insulin fibril assembly are dominated fragmentation (Reaction 2). The nucleation event becomes important when the system volume is small (pico-liter scale) [127]. The working experimental volumes are much larger compared to such system sizes (100 μL). Therefore, fragmentation is the dominant process in describing the kinetics of assembly and an explicit account of nucleation is not required to capture the features of the kinetic experimental data.

We find that our simulations best fit our experimental data (in particular for the scaling of lag time with protein concentration) if all simulations are initiated with a single dimer present ($f_2 = 1$ at time zero). In our simulations, we track explicitly the growth and fragmentation of individual fibrils. Each simulation data point shown represents the average (and standard deviation) of 150 replicate runs. Typical run-times are on the order of 100,000 seconds. The values of k_+

and k_f were obtained by fitting our experimental data to the model of Knowles, *et al.* [119], in the low concentration regime ($c_p \leq c_T$; k_+=5x10$^4 s^{-1}$; k_f=1x10$^{-8} s^{-1}$) and also for the lowest NaCl concentration (0.1M) (k_+=1x10$^6 s^{-1}$; k_f=1x10$^{-8} s^{-1}$). These values of the rate constants obtained from fits to the experimental data were then used as the parameter values in the simulations outlined above.

Chapter 3

Insulin Amyloid Fibril Formation: Mechanistic & Environmental Control of the Prevalence & Lifetime of Amyloid Oligomers

3.1 Introduction

A wide variety of peptides and proteins form self-assembled filamentous structures known as amyloid fibrils. A universal feature of all amyloid aggregates, regardless of the amino acid sequence or native conformation of the starting peptide or protein, is the formation of cross-β sheet structure [159, 55, 114, 160]. Amyloid fibrils are most commonly associated with human diseases such as Alzheimers Disease and type-2 diabetes [19], however, they are also garnering interest in biomaterial applications [86, 75, 161, 79, 77, 162, 81, 163].

Understanding the molecular mechanism(s) leading to amyloid fibril assembly is a key focus of current research. Mechanistic models of amyloid fibril self-assembly are often tested by monitoring the kinetics of formation using absorbance or fluorescence spectroscopy. The kinetics of amyloid fibril formation are typified by a lag phase followed by the onset of rapid growth. This has been attributed to the formation of a critical nucleus via primary nucleation [49](which

may be very rapid) followed by growth via fragmentation, heterogeneous nucleation or branching [115]. This classical picture has been extended to include subtle effects of protein concentration [164], multi-stage kinetics [165, 166, 167, 168], and models which include competition between multiple pathways [169].

Recent research has focused on the role of fragmentation in fibril formation kinetics. Xue *et al.* [170] first noted that fragmentation is central to explaining the kinetics of β2-microglobulin aggregation, which cannot be described by a straightforward nucleated assembly model. More recently, analytical results show that several important features of fibril growth kinetics can be explained by a model in which fragmentation is the dominant process in fibril assembly [119, 120, 121, 122]. The classical nucleation picture [116], in which the lag time, i.e. the duration of the lag phase, is determined by the time it takes to form a thermodynamically unfavorable critical nucleus, predicts that the lag time should scale with protein concentration as $\tau_{lag} \sim c_p^{-\gamma_c}$, where $\gamma_c = \frac{n_c+1}{2}$ [115]. Since the critical nucleus size $n_c > 1$, in classical models the scaling exponent will always be greater than unity. In contrast, a model dominated by fragmentation, in which fibril fragmentation acts as a positive feedback mechanism producing new ends that can then grow, will exhibit a scaling exponent $\gamma_{frag} \sim \frac{1}{2}$ [119]. This prediction is largely consistent with experimental observation of many known amyloidogenic protein systems. However, fragmentation-dominated models predict exponentially decaying fibril length distributions, while the majority of experimental measurements show distributions peaked at intermediate fibril length. [171, 172, 173].

Here I resolve this apparent contradiction by presenting a qualitatively new model for amyloid fibril assembly. Using a series of experiments on bovine insulin, combined with stochastic simulations, I find evidence to support the existence of a critical fibril mass concentration (CFC) above which fibrils undergo a structural transition that suppresses fragmentation. I attribute the onset of arrested fragmentation to a transition of the amyloid fibril solution from a viscous fluid to a viscoelastic gel. Our model predicts the formation of distinct fibril length distributions for protein concentrations above and below a critical threshold and I confirm these by experiment. Moreover I show that the concept of a structural transition that arrests the effects of fragmentation enables us to successfully predict the influence of environmental conditions on the fibril

assembly pathway. Our results suggest a possible explanation for the observed variation in amyloid disease progression among apparently similar individuals: since the lengths of fibril formed depend critically on the protein concentration and are influenced by environmental conditions, in some individuals the intra- and/or inter-cellular environment may favor the formation of predominantly short and possibly toxic fibrillar species, while in other individuals conditions may result in benign fibril loads. This work has recently been published in Nature Communications.

3.2 Results

3.2.1 ThT Fluorescence Assays

The kinetics of bovine insulin amyloid fibril formation were monitored using ThT fluorescence in 96-well microwell plates coated with a hydrophilic polymer (§2.2.2). At least 136 replicates were performed for each set of conditions, resulting in highly accurate results. The inset to Fig. 3.1 shows typical results from a ThT fluorescence assay of a single microwell plate. I am interested in primarily two kinetic parameters obtainable from the kinetic traces (averaged over replicates): the growth rate, k_{gr}, and the lag time τ_{lag}. The growth rate,k_{gr}, is found by fitting a linear function to the points of maximal growth. In turn, the lag time is obtained by extrapolating this fit line to time-axis. This is illustrated in Fig. 3.1. There are many definitions of the lag time in the literature [168, 39, 123] but this definition was adopted to compare results with [119].

It should be noted how reproducible the kinetics are between wells during a given experiment (Fig 3.1, inset). Previous reports have shown the lag times measured for identical protein samples to vary by as much as hours, and this has been attributed to the stochastic nature of the nucleation process [123, 10] (see §1.4.3): in our experiments, due to careful control of experimental conditions, the variation in the lag time has been reduced to minutes (see Appendix B for lag time distributions). Our method thus enables us to identify subtle changes in insulin amyloid fibril growth kinetics upon changing the experimental parameters. While considerably reduced, itt should be noted that there still exists some variation between replicate experiments. Such variation is not completely understood, however, a proportion may be due systematic errors introduced by pipetting

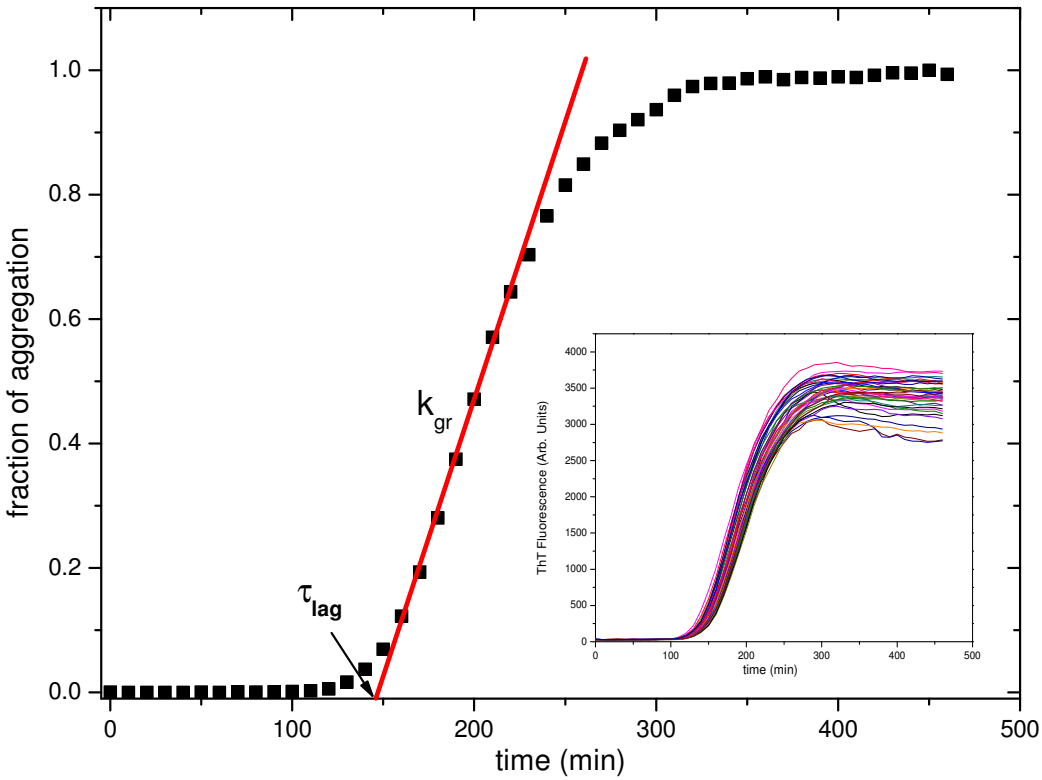

Figure 3.1: **Representative kinetic trace for insulin amyloid fibril formation as monitored by ThT fluorescence.** The plot shows a kinetic trace that has been normalized to its maximum. A linear function is fit through the region of maximum growth. The growth rate is defined as the slope of the fit curve. The lag time is determined by extrapolating the fit curve to the time-axis. Inset shows well-to-well variation in the kinetics of insulin amyloid fibril formation. Replicate experimental kinetics curves show the variability typical of replicate experiments within a microwell plate, for an insulin concentration of 1.5 mg/ml.

errors, or imperfectly homogeneous heating of the samples (see Appendix A).

3.2.1.1 Kinetics of unagitated samples

I performed a set of kinetic measurements without any agitation of the plate during the time course of fibril assembly. Initial insulin concentrations ranged from 0.1 mg/ml up to 4 mg/ml. A fluorescence reading was recorded at 10 minute intervals for all experiments. A plot of the mean lag time as a function of insulin concentration is shown in Fig. 3.2. The red solid line in Fig. 3.2 is a fit of the data to a power law function of the form $y = ax^b$. I find the scaling

exponent, γ, is -0.37 $\pm$ 0.02. Since $\gamma < 1$, the data are broadly consistent with a model dominated by fragmentation [119].

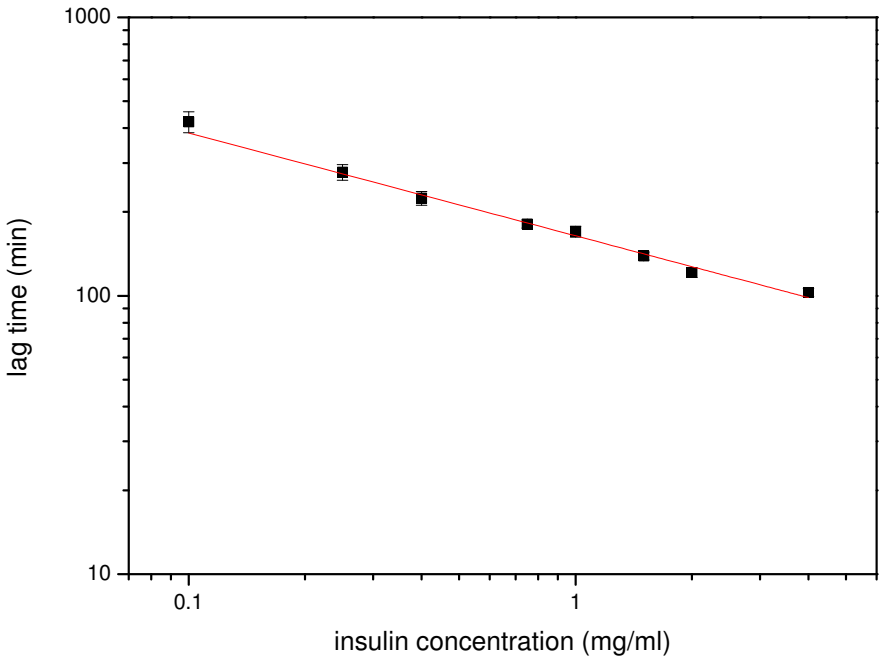

Figure 3.2: **Lag time vs. insulin concentration for unagitated samples.** Each data point in the plot consists of 140-200 individual experiments carried out in several microwell plates. The plates were not agitated during the time course of fibril assembly. Lag times are plotted as a function of insulin concentration on a log-log plot. The data is fit with a generic power law function $y = ax^b$. The scaling exponent γ, is -0.37 $\pm$ 0.02. Data points show mean $\pm$ s.d.

I am also interested in how the rate of growth of ThT fluorescence, which I will call k_{fluo}, varies with insulin concentration. This trend is shown in Fig. 3.3. There are two interesting features to this plot. Firstly, in a 'low' concentration regime (from 0.1-0.4 mg/ml), k_{fluo} increases linearly with concentration. Secondly, however, at $\sim$ 0.75 mg/ml this trend becomes non-linear. One can also look at how the mean maximal ThT fluorescence varies as a function of insulin concentration (Fig. 3.4). Comparing Fig. 3.3 with Fig. 3.4 is becomes apparent they exhibit very similar trends. Indeed, both show marked increases in k_{fluo} and the mean maximal fluorescence above 0.75 mg/ml. This effect is not caused by a change in the total fibril mass; pelleting the aggregated material by centrifugation

and measuring the remaining protein in the supernatant by UV-Vis spectroscopy confirmed that more than 99% of the total insulin present is in the pellet regardless of the initial protein concentration (see §2.2.3).

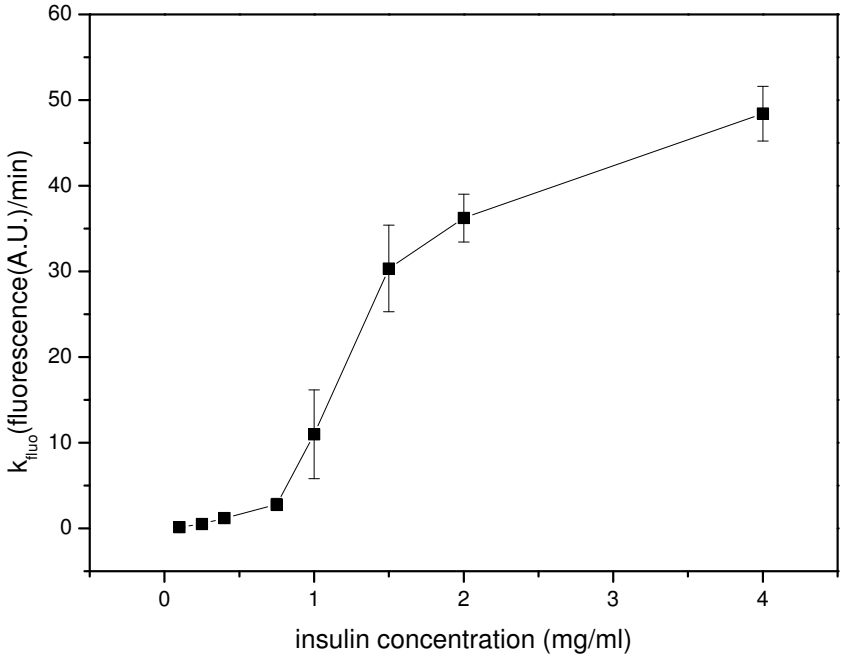

Figure 3.3: **Maximum rate of fluorescence increase vs. insulin concentration for unagitated samples.** The rate of maximum fluorescence increase, k_{fluo}, was determined as shown in Fig. 3.1. Plot shows k_{fluo} as a function of insulin concentration. From 0.1-0.75 mg/ml there is a linear trend; between 0.75 and 1 mg/ml a sharp increase in fluorescence is observed. Data points show mean ± s.d.

3.2.1.2 Kinetics of agitated samples

Since the model of [119] relies on fragmentation as being the key driving mechanism in fibril assembly I wished to further test this idea using our experimental methods. I performed a second set of experiments, identical to those performed in the previous section, but now with addition of plate agitation during fibril assembly. The plates were shaken at 600 rpm in a double orbital motion for two minutes with a total cycle time between fluorescence readings of

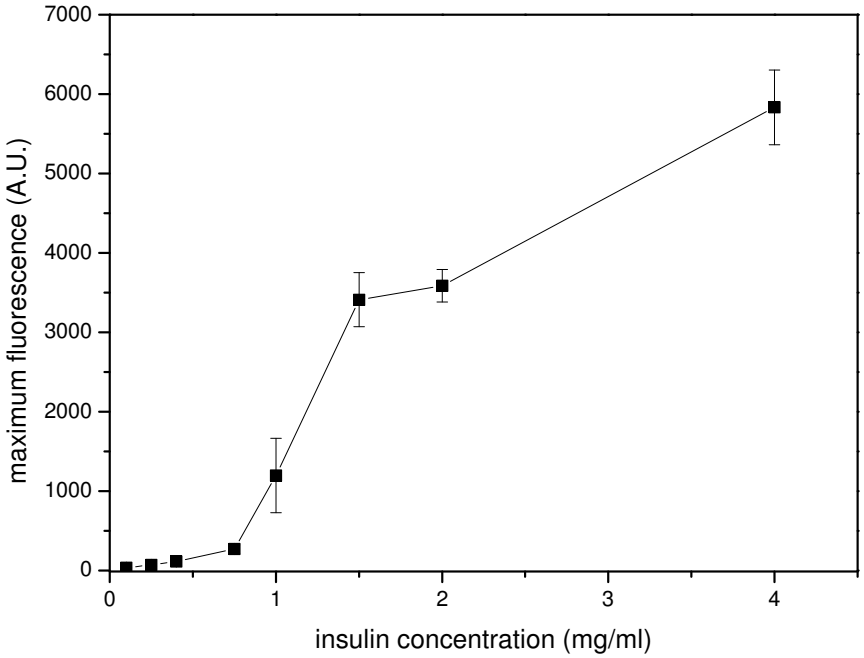

Figure 3.4: **Mean maximum fluorescence vs. insulin concentration for unagitated samples.** The maximum fluorescence values show similar characterstics as in Fig. 3.3: a sharp increase in the maximum fluorescence value between 0.75 and 1 mg/ml. Data points show mean ± s.d.

10 minutes (see §2.2.2.1 and Appendix A). I performed this additional shaking step because Xue and co-workers showed that this results in fibril fragmentation [170]. Again, I can observe at how the lag time scales with protein concentration (Fig. 3.5). In the presence of agitation I find a scaling exponent of -0.50 ± 0.03. Indeed, this is in excellent agreement with the prediction of the model of Knowles *et al.*.

The trends in k_{fluo} and the mean maximal fluorescence as a function of concentration are plotted in Fig. 3.6 and Fig. 3.7. Comparing the agitated results to the unagitated ones, I find that the growth rates and fluorescence are greater, but the overall trends are very similar. One can surmise that agitating the samples results in a more efficient means of fragmenting the fibrils, but the underlying mechanisms of fibril assembly, between the two data sets, are the same. For a compilation of all the kinetic parameters pertaining to the experimental data

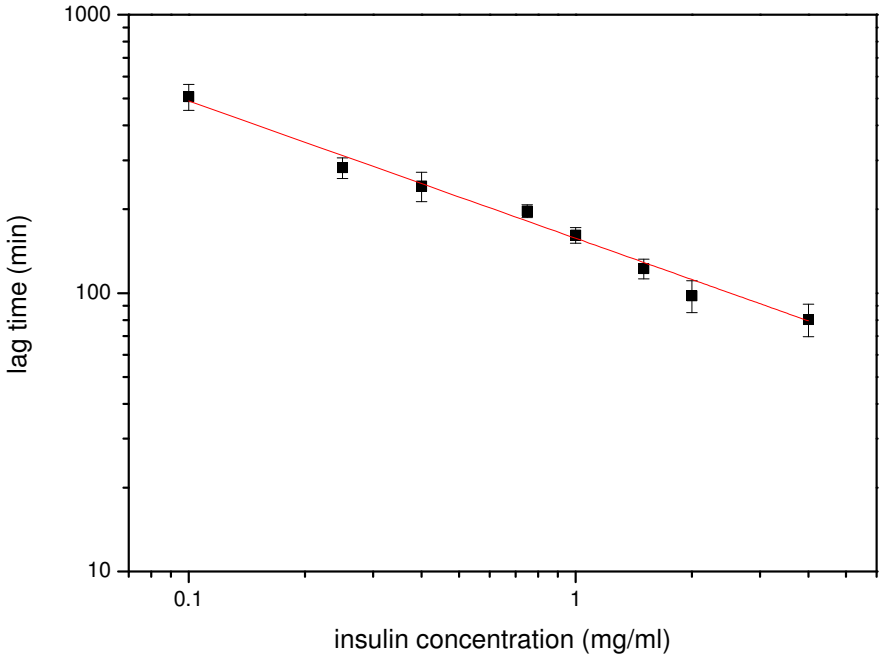

Figure 3.5: **Lag time vs. insulin concentration for agitated samples.** The plates were agitated in a double orbital motion at 600 rpm for 2 minutes for all experiments. Lag times are plotted as a function of insulin concentration on a log-log plot. The data is fit with a generic power law function $y = ax^b$. The scaling exponent γ, is -0.50 $\pm$ 0.03. Data points show mean $\pm$ s.d.

of both agitated and unagitated samples, see Appendix C.

3.2.1.3 Emergence of two concentration regimes

What becomes apparent by looking at the results above is that: 1) a model of fragmentation dominated fibril assembly provides a fair description of the kinetics of insulin amyloid fibril formation (particularly in the lag time) and 2) the emergence of two apparently distinct kinetic regimes as a function of initial insulin concentration. As is shown in Fig. 3.3 and Fig. 3.6, k_{fluo} is linear in a concentration range from 0.1-0.4 mg/ml but becomes non-linear at concentrations above 0.75 mg/ml. Furthermore, since the mean maximal fluorescence trends similarly to k_{fluo} (i.e. in a non-trivial way) it would be informative to refer to a growth rate which is independent of the fluorescence growth rate. To decouple

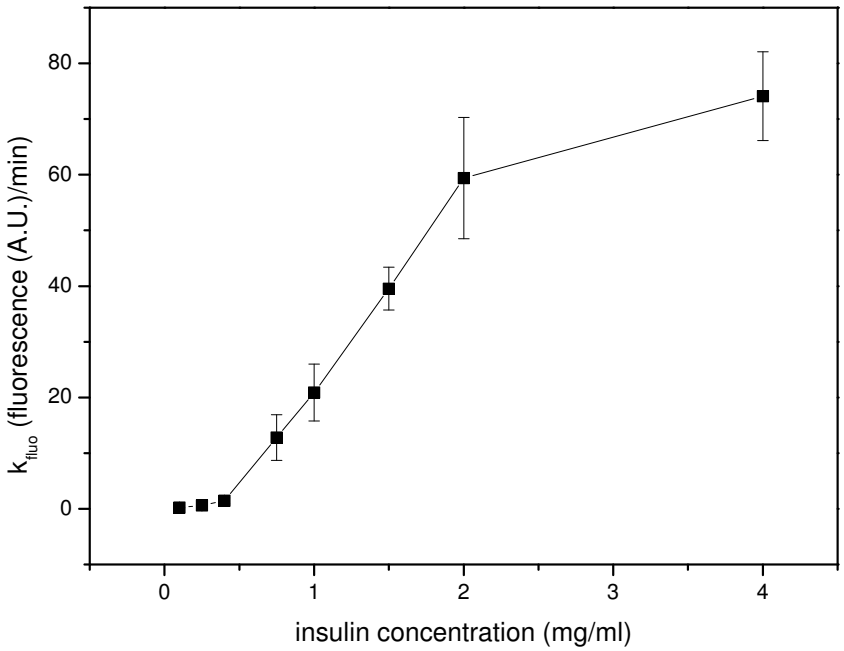

Figure 3.6: **Maximum rate of fluorescence increase vs. insulin concentration for agitated samples.** The rate of maximum fluorescence increase, k_{fluo}, was determined as shown in Fig. 3.1. Plot shows k_{fluo} as a function of insulin concentration for agitated samples. I find a non-linear trend emerges ~ 0.75 mg/ml. Data points show mean $\pm$ s.d.

the growth rate measurements from the changes in maximal fluorescence, the growth curves were normalized to their maximal values and the maximal growth rate, k_{gr}, was extracted. To ensure that the fluorescence growth rates that are measured are not artifacts of some fluorescence anomaly we simultaneously measured absorbance. Fig. 3.8 shows k_{fluo} plotted against k_{abs}. What is clear is that the measured fluorescence rate is correlated well to the measured growth rate from absorbance. This gives confidence that $k_{gr}$is not influenced by fluorescence artifacts.

The emergence of these two concentration regimes becomes all the more apparent when k_{gr} is plotted as a function of insulin concentration (Fig. 3.9). It is clear that for protein concentrations greater than 0.75 mg/ml the growth rate becomes nearly independent, if not slightly decreasing, with insulin concentration.

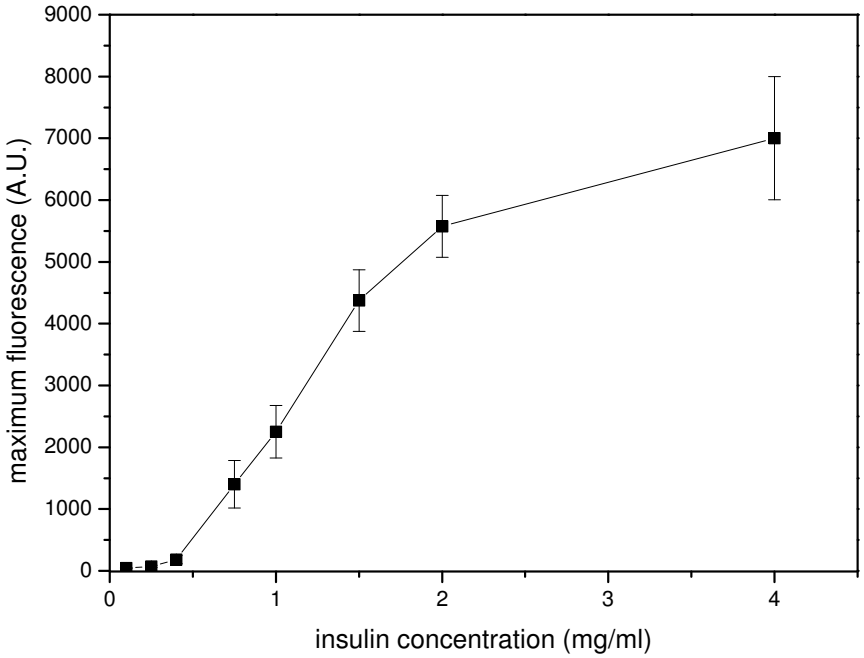

Figure 3.7: **Mean maximum fluorescence vs. insulin concentration for unagitated samples.** The maximum fluorescence values show similar characterstics as in Fig. 3.6: a sharp increase in the maximum fluorescence value ~ 0.75 mg/ml. Data points show mean $\pm$ s.d.

Perhaps more importantly, Fig. 3.9 also shows the predicted growth rate from the model of [119] (blue triangles). This plot was obtained by performing simulations based upon the model of [119] using kinetic parameters obtained from our experiments at low insulin concentrations. What is strikingly clear is that at a concentration of 0.75 mg/ml there is a marked deviation between the experimentally obtained growth rates and those predicted by the model. The insulin concentration at which we observe this deviation from the model is defined as 'threshold' concentration, or c_T.

3.2.2 Circular Dichroism of Insulin Amyloid Fibrils

It is conceivable that the sudden order-of-magnitude increase in fluorescence (Fig. 3.4, Fig. 3.7) is due to structural or morphological differences (presumably affecting the binding of the fluorophore) between fibrils formed at 'high' and

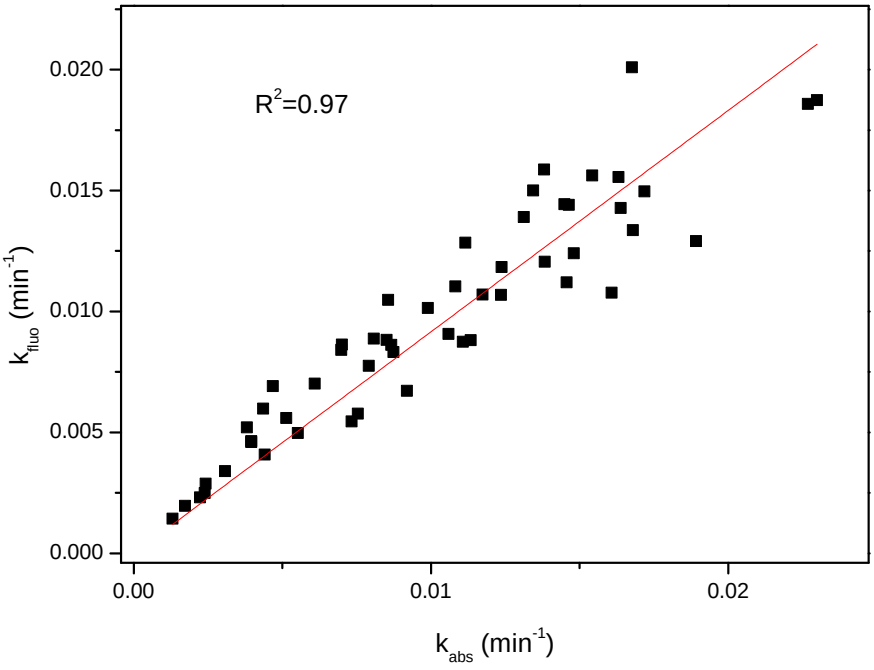

Figure 3.8: k_{fluo} as a function of k_{abs}. We measured the ThT fluorescence and absorbance during insulin amyloid fibril formation for 1 mg/ml insulin. We find that the measured rates in both techniques are well correlated.

'low' insulin concentrations. Moreover, the action of agitating the samples may have a demonstrable affect upon the structural characteristics of fibrils formed under those conditions since the fluorescence increase is even greater in those experiments. It is known that the structural characteristics of amyloid fibrils are important in the binding of ThT [47, 48, 46]. For example, it has been shown that glucagon fibrils formed at 'high' ($>$ 1 mg/ml) and 'low' ($<$ 1 mg/ml) protein concentrations demonstrate markedly different structural characteristics and ThT fluorescence emission signals [58]. In order to investigate whether there are measurable structural differences between fibrils formed at different concentrations in our experiments I measured circular dichroism (CD) spectra of insulin amyloid fibrils prepared at concentrations between 0.25 mg/ml and 4 mg/ml, i.e., above and below c_T (see §2.2.7.1). I find that the spectra of the fibrils prepared at high (1-4 mg/ml) and low (0.25-0.75 mg/ml) protein concentrations are qualitatively different(Fig. 3.10). Indeed, there is a striking transition in

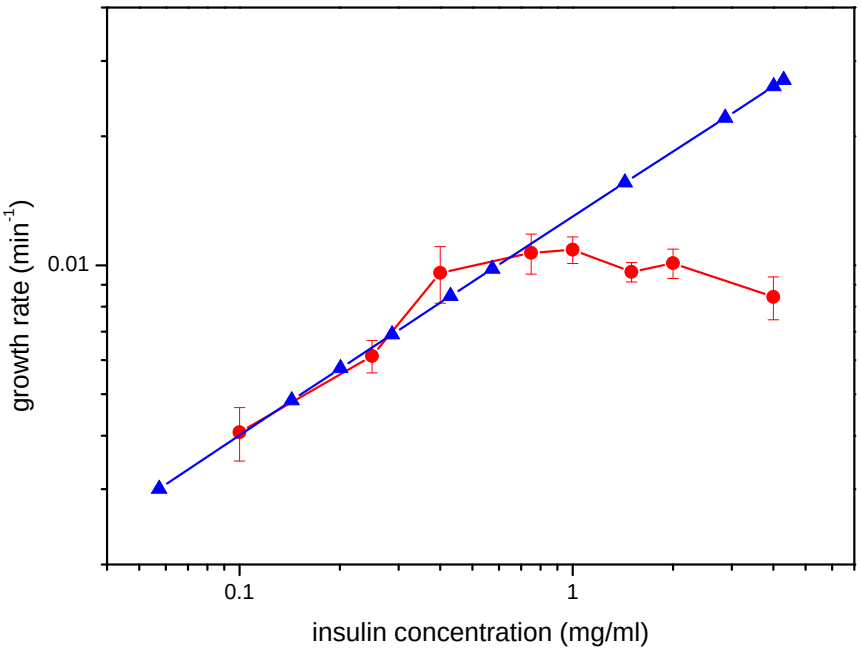

Figure 3.9: **Normalized growth rate as a function of insulin concentration: predicted and experimental** Kinetic curves were normalized to their maxima and the normalized growth rate, k_{gr} were extracted the same manner as determining k_{fluo}. There is a linear increase in k_{gr} between 0.1 and 0.4 mg/ml; for concentrations greater than 0.4 mg/ml k_{gr} reaches a plateau (red circles). For comparison, the predicted k_{gr} from a fragmentation dominated model is shown (blue triangles). Data points show mean ± s.d.

the circular dichroism signal between 0.75 and 1.5 mg/ml, consistent with the transition seen in our fluorescence measurements. At low concentrations (<0.75 mg/ml) I attribute the gradual decrease in the CD spectra to changes to fibrillar association, that is, from single filaments to double [7]. Above 1 mg/ml I ascribe the contrasting increase in signal to a structural transition similar to that observed for small aromatic molecules on changing from the solution to the crystalline state [8].

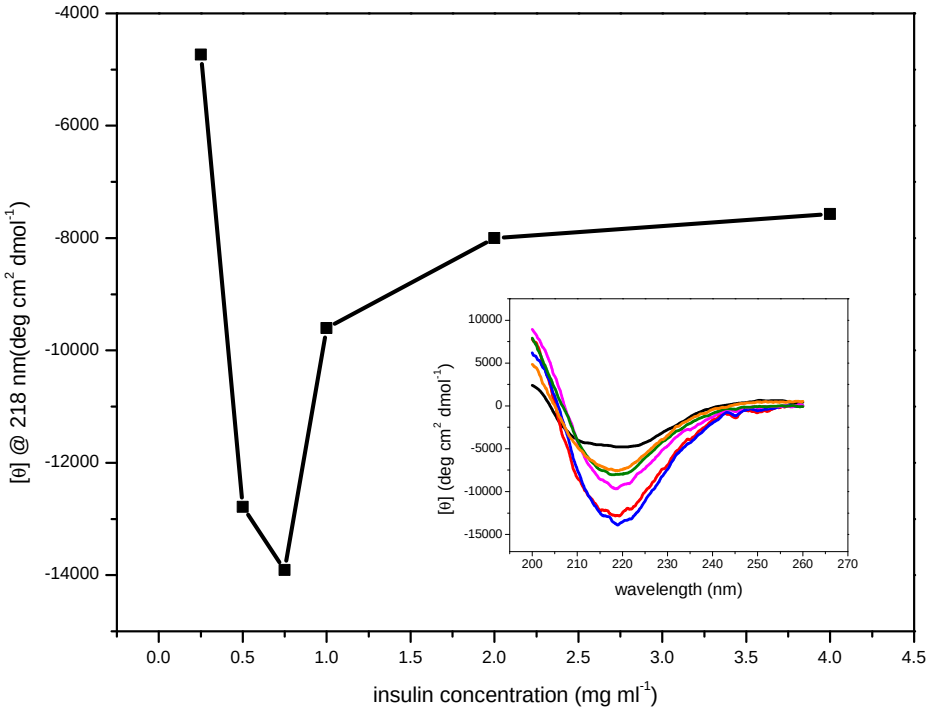

Figure 3.10: **CD spectra of insulin amyloid fibrils formed at initial protein concentrations above and below** c_T. Plotted are the molar ellipticities at the minimum (218 nm) of the CD spectra (inset) as a function of the insulin concentration. The inset shows the CD spectra for insulin amyloid fibrils formed from 0.25 mg ml^{-1} (black), 0.5 mg ml^{-1} (red), 0.75 mg ml^{-1} (blue), 1 mg ml^{-1} (pink), 2 mg ml^{-1} (green), 4 mg ml^{-1} (orange). There is a striking cross-over between regimes reflected in the CD signal above and below 1 mg ml^{-1} insulin coincident with where the kinetic transition occurs. At low concentrations (<0.75 mg ml^{-1}) I attribute the gradual decrease in the CD spectra to changes in the filament organization with increasing protein concentration (i.e. from single filaments to double [7]). Above 1 mg ml^{-1} I ascribe the contrasting increase in signal to a structural transition similar to that observed for small aromatic molecules on changing from the solution to the crystalline state [8]

3.2.3 Microrheology of insulin amyloid fibrils

To better understand this deviation from the predicted growth behavior I performed bead tracking microrheology on amyloid fibril samples obtained from a range of insulin concentrations that had been incubated for 24 hours (see §2.2.6.3). I find a transition from a liquid to a highly viscous fluid state for fibrils

formed at protein concentrations above 0.75 mg/ml (Fig. 3.11; see §2.2.5). Taken together, these findings lead us to propose that there exists some critical fibril mass concentration, or 'CFC', above which fibril-fibril assembly, bundling, and/or gelation dominates [137, 174] and as a consequence fragmentation is suppressed.

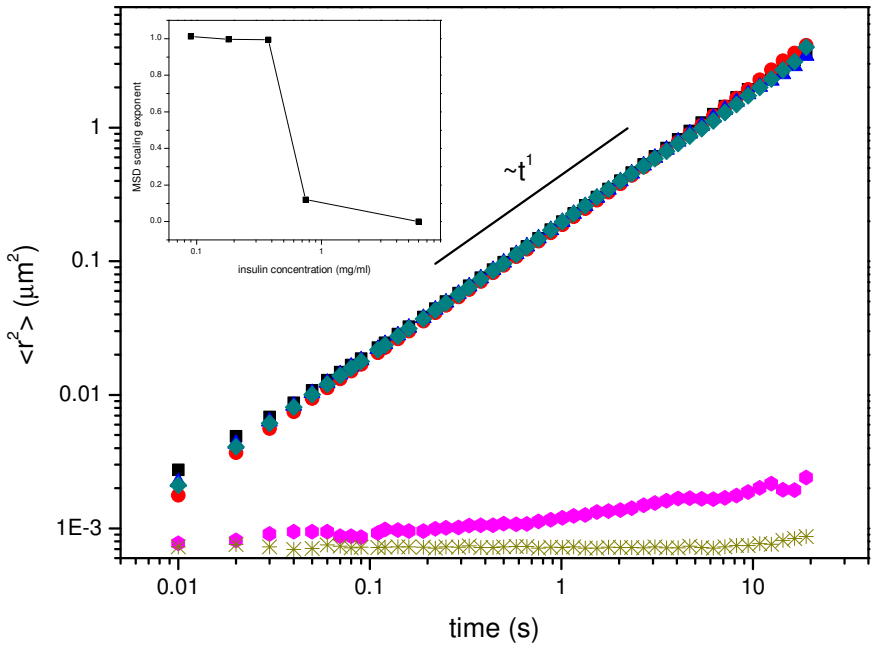

Figure 3.11: **2D MSD of insulin amyloid fibril solutions shows a transition to a gel-state at 0.75 mg/ml.** The mean square displacement (MSD) of 1 μm beads are shown for a range of insulin concentrations: buffer control (squares), 0.09 mg/ml (circles), 0.18 mg/ml (triangles), 0.375 mg/ml (diamonds), 0.75 mg/ml (hexagons), 6 mg/ml (stars). For particles undergoing Brownian motion, the MSD scales as $\langle r^2 \rangle \sim t^\alpha$, where α is the scaling exponent. Black line indicates α=1. For concentrations below 0.75 mg/ml the insulin amyloid fibril solutions exhibit the characteristics of a viscous fluid. At 0.75 mg/ml, the MSD scaling exponent drops from 1 to ~0.2 indicating a transition to a highly viscous fluid. Inset shows the MSD scaling exponents as a function of insulin concentration.

3.2.4 Stochastic Simulations of Amyloid Fibril Assembly

To investigate the idea that above a critical fibril concentration (CFC) fragmentation becomes arrested we performed stochastic simulations of fibril assembly. We performed stochastic simulations of a fragmentation-dominated model of fibril growth (§2.2.9). The kinetic parameters used in the simulation were obtained by fitting our experimental data at low protein concentrations to analytical results derived by Knowles *et al.*[119]. This resulted in elongation and fragmentation rates of $5\text{x}10^4 M^{-1}s^{-1}$ and $1\text{x}10^{-8}s^{-1}$, respectively (for details see §2.2.9). For low protein concentrations we find that the values of k_{gr} predicted by these simulations are in good agreement with our experiments. For high protein concentrations, however, our simulations of the fragmentation-dominated model fail to match our experimental results (c.f. Fig. 3.9). To test our hypothesis that a structural transition is involved we modified our simulations so that fragmentation is arrested when the total fibril mass concentration exceeds a critical value (the CFC). In our simulations, a value for the CFC of 0.4 mg/ml reproduces our experimental results for k_{gr} very well over the entire protein concentration range. We re-plot Fig. 3.9 now with the addition of our simulations with a CFC (Fig. 3.12).

A natural question is how does the CFC value relate to the observed 'threshold' concentration, c_T? Intuitively, for the CFC to have an effect on k_{gr}, the CFC must be reached before k_{gr}. The maximum growth rate usually occurs at approximately 50% of the total aggregation (c.f. Fig. 3.1). Therefore, the CFC must always be *at most* half the total concentration of initial protein. Therefore, the threshold concentration c_T must be twice the CFC value. The observed transition occurs $\sim$ 0.75 mg/ml which corresponds to a CFC of $\sim$0.375 mg/ml. As stated above, we find a CFC of 0.4 mg/ml best matches our experimental data; this corresponds well with the anticipated value.

3.2.5 Transmission Electron Microscopy of Insulin Amyloid Fibrils & Length Distributions

A strong prediction which comes out of this model is that the final fibril length distributions in the long time limit should be markedly different at high and low initial protein concentrations. For initial protein concentrations less than c_T,

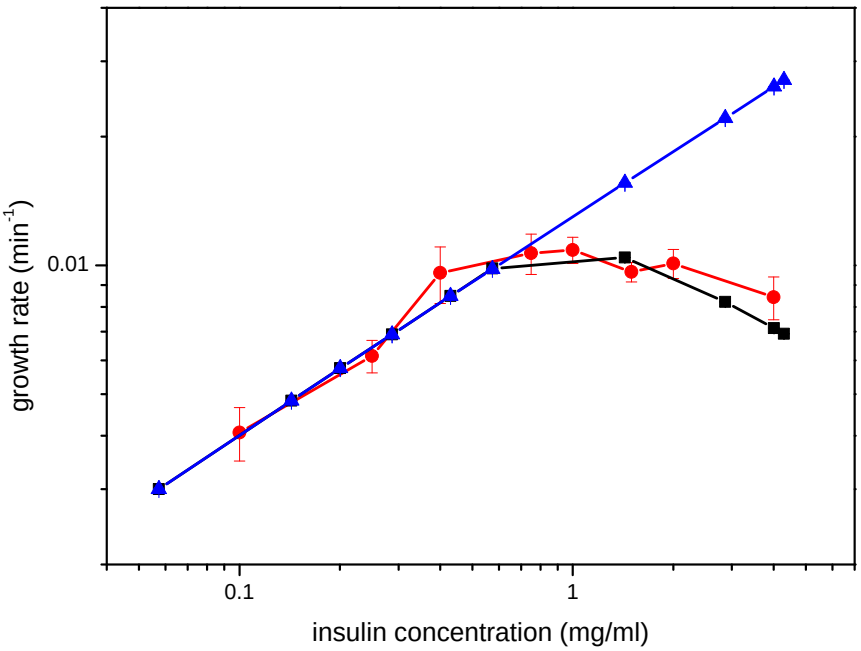

Figure 3.12: k_{gr} **as a function of insulin concentration: comparing experiment and simulation.** With the addition of the CFC into the simulation, we find a close agreement betweens simulation (black squares) and experiment (red circles). For comparison, the prediction from the model is shown without a CFC (blue triangles).

fibrils continue to fragment throughout the simulation, and in the long-time limit the fibril length distribution is exponential and dominated by short fibrils, as predicted by fragmentation-dominated models (Fig. 3.13(B)). In contrast, when the initial protein concentration is greater than c_T, the mass concentration of fibrils exceeds the CFC at some time during the exponential growth phase so that fragmentation ceases. This results in a log-normal fibril length distribution with a clear peak at intermediate fibril lengths, in agreement with many experimental observations (Fig. 3.13(A)). Thus, when the protein concentration c_p is greater than c_T the length distribution is prevented from collapsing due to a structural transition that suppresses the action of fragmentation.

To test these predictions I used transmission electron microscopy (TEM) to obtain length distributions for amyloid fibrils formed at initial protein

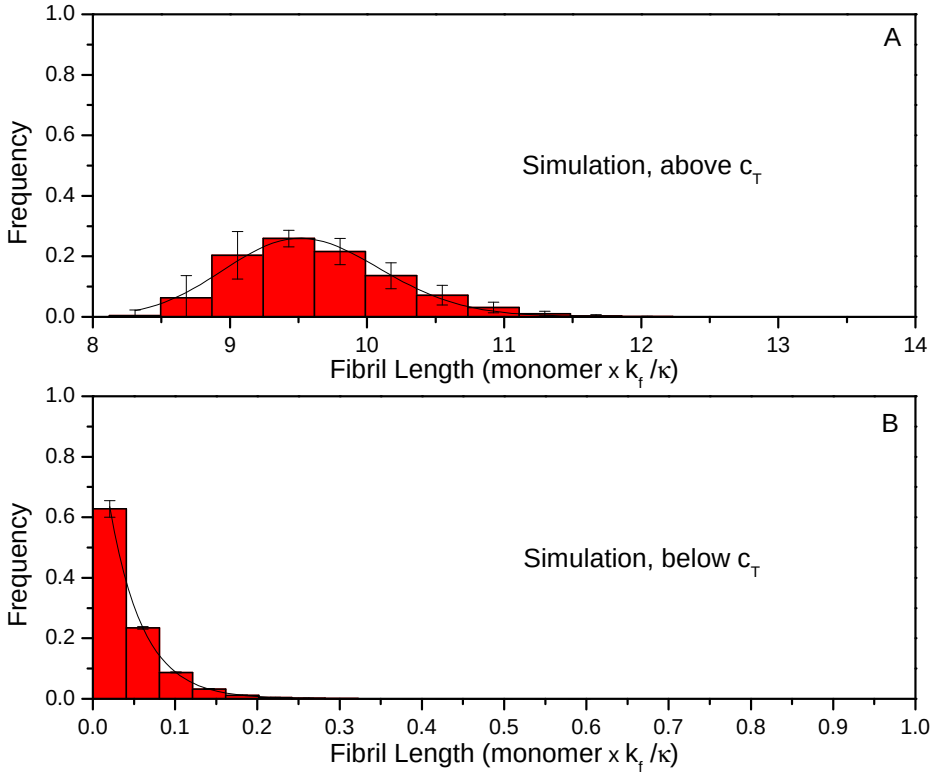

Figure 3.13: **Fibril length distributions obtained by simulation** Fibril length distributions obtained by simulation (**A**) above c_T (4 mg/ml) and (**B**) below c_T (0.2 mg/ml). We express the fibril length in terms of the composite parameter κ/k_f . Our simulations do not make absolute predictions of fibril length because the choice of values of k_+ and k_f is not strongly constrained (§2.2.9), whereas the shape of the length distribution (plotted as a function of monomer units $\times$ k_f/κ) is independent of this choice. The length distribution in (**A**) is best described by by a log-normal distribution; (**B**) is well-fitted by an exponential distribution (black fit lines).

concentrations above and below our predicted c_T of 0.75 mg/ml (4 mg/ml and 0.2 mg/ml, respectively. See §2.2.5.1). A representative image is shown in Fig. 3.14. Using images acquired from TEM, lengths were determined using ImageJ software. The resulting length distributions are shown Fig. 3.15. Our experimental results are in close correspondence with our simulation predictions: indeed, I observe an exponential distribution of fibril lengths when the protein concentration is below c_T and a peaked distribution for protein concentration above c_T (Figs. 2A,C). Many other studies have reported peaked fibril length distributions [171, 172, 173]. I hypothesize that a structural transition may also

be operating in these systems, and that these experiments may have been carried out in the high protein concentration regime where $c_p > c_T$.

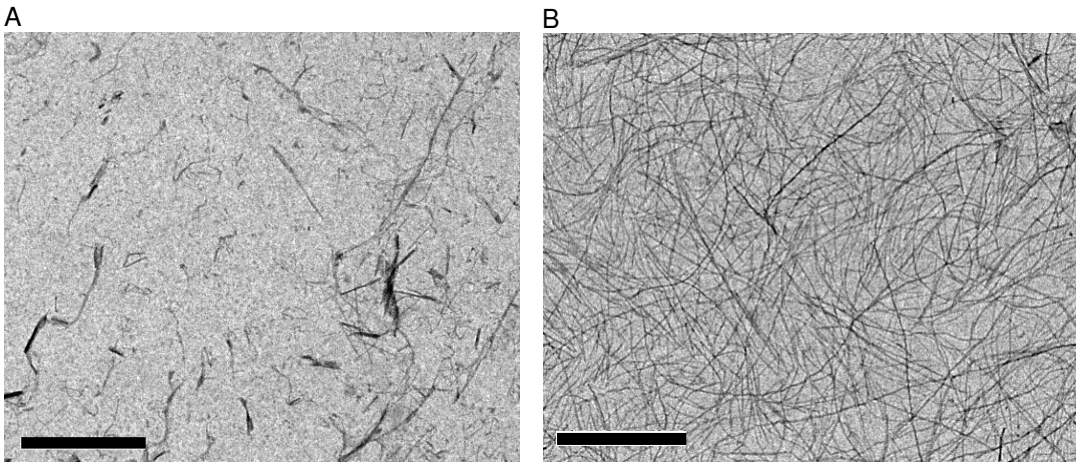

Figure 3.14: **Representative TEM images of insulin amyloid fibrils formed above and below the CFC**. Images of insulin amyloid fibrils formed from initial insulin concentrations of (**A**) 0.2 and (**B**) 4 mg/ml. See Fig. 3.15 for corresponding length distributions. Scale bar is 1 μm.

3.2.6 Addition of NaCl & Effect on the Kinetics

Our model further allows us to rationalize non-trivial changes in the kinetic curves when the environmental conditions are altered. It is well known that the addition of salt screens electrostatic charge, thereby decreasing the Debye length and influencing protein-protein interactions [175]. As a model environmental perturbation, I therefore repeated our kinetic experiments in the presence of increasing concentrations of NaCl.

Fig. 3.17 shows k_{gr} as a function of insulin concentration for a range of NaCl concentrations. It is clear that adding NaCl results in complicated and non-monotonic trends in the kinetics of fibril growth. While these trends are not readily captured by variation of the parameters in fragmentation-dominated models, our simulations in which a structural transition is included are in excellent agreement with the data if one makes the simple assumption that salt affects both the elongation rate (k_+) and the CFC, but not the fragmentation rate. The results of our simulations are shown in Fig. 3.18. Firstly, the increase in growth rates between 0 M and 0.1 M NaCl is consistent with a $\sim$ 20 fold

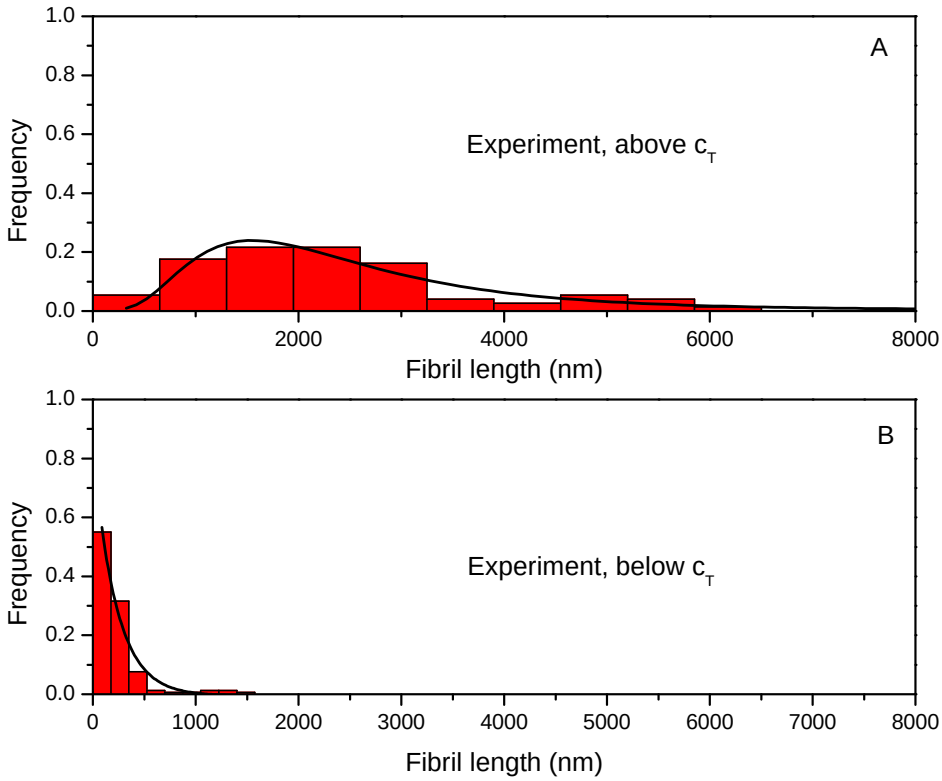

Figure 3.15: **Fibril length distributions extracted from TEM** (**A**) above c_T (4 mg/ml, N=159, bars) and (**B**) below c_T (0.2 mg/ml, N=75, bars). As predicted by the model, fibrils formed at protein concentrations below the CFC are significantly shorter (mean length of 231.5 ± 230.4 nm) and the distribution is exponential (black curve). Length distributions above c_T show a log-normal distribution (black curve) with a mean fibril length of 2441.5 ± 1590.3 nm (mean ± s.d.). Experimental and simulation data points show mean ± s.d.

enhancement of the fibril elongation rate to $1\text{x}10^6 s^{-1}$. Secondly, we capture the complex, salt-dependent and non-monotonic trends of the growth rate with protein concentration in Fig. 3.17 by allowing the CFC to decrease with salt concentrations above 0.1M. It should be emphasized that the changes in the simulated curves at 0.25-0.5 M NaCl arise only from a change in the CFC; the elongation rate is held constant at $1\text{x}10^6 s^{-1}$. Further increasing the elongation rate does not account for the complex behavior observed in Fig. 3.17; the presence of a structural transition in the model is essential in order to explain the experimental data. Further evidence to support the hypothesis that salt is shifting the CFC to lower values is found in Fig. 3.16. I find that the non-linearity in the ThT fluorescence occurs at lower and lower insulin concentrations as the

NaCl concentration is increased. This is evidence for the fact that the CFC value decreases with increasing NaCl concentrations.

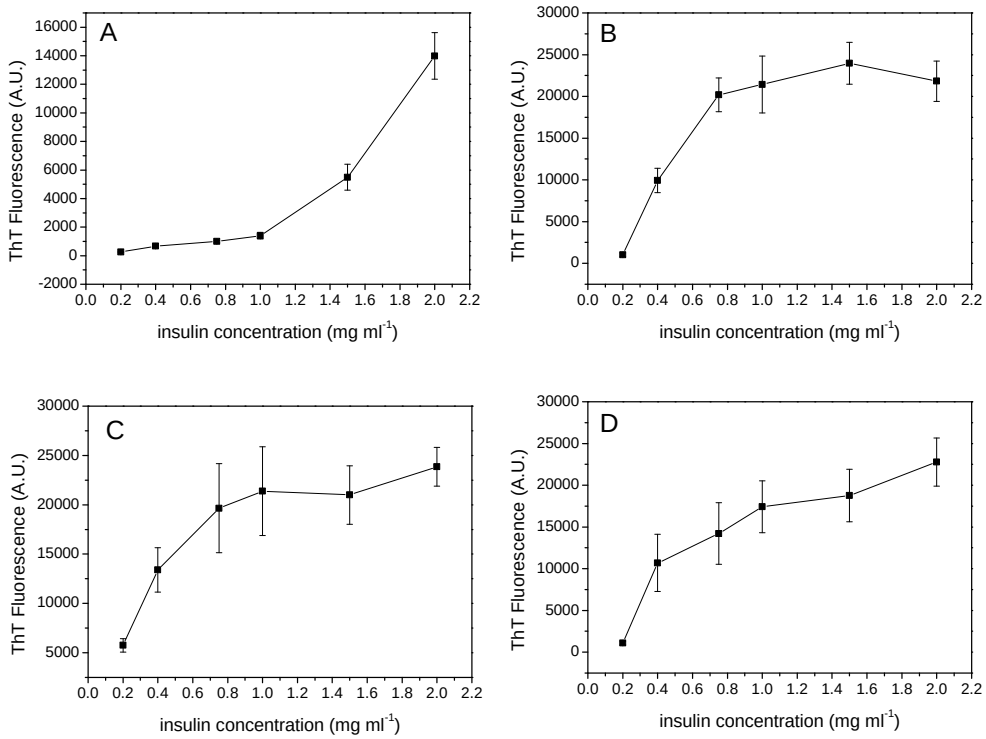

Figure 3.16: **The trends in maximum ThT fluorescence as a function of concentration and NaCl.** Plotted are the mean maximum ThT fluorescence values, as a function of insulin concentration, for (**A**) 0.1 M, (**B**) 0.24 M, (**C**) 0.34 M, and (**D**) 0.49 M NaCl concentrations. I find that the concentration at which the non-linearity in the ThT fluorescence occurs at decreasing insulin concentrations as the NaCl concentrations is increased. This observation can be explained by positing that the CFC is shifting to lower concentrations.

It should be noted that the plateau in the elongation rate with increasing salt concentration is not the result of reaching the diffusion limit. A simple calculation yields a diffusion-limited growth rate on the order of $1\text{x}10^{9} M^{-1} s^{-1}$ which is several orders of magnitude below our experimentally obtained growth rate. The value obtained above can be understood by positing that the addition of monomer to the growing fibril end becomes limited by the time scale for the monomer to adopt the correct conformation to allow the next monomer to attach [176, 177].

Our structural transition model also helps us to understand how the lag time,

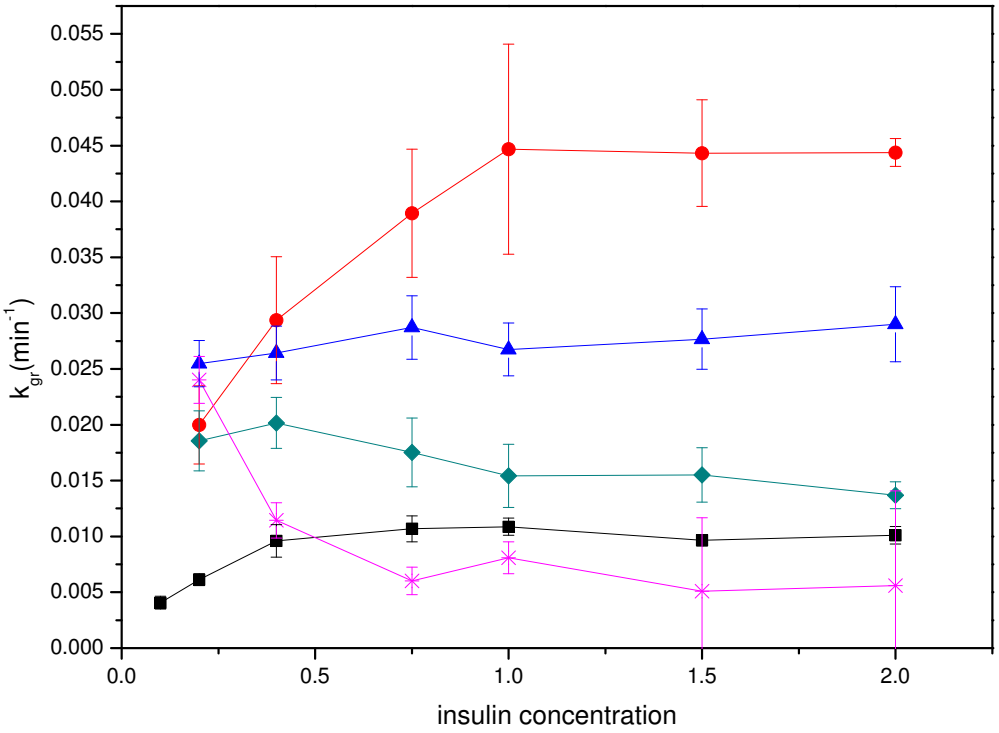

Figure 3.17: **Experimental k_{gr} as a function of insulin concentration with the addition of NaCl**. Growth rates obtained from normalized traces plotted as a function of insulin concentration for NaCl concentrations of 0M (black squares), 0.1M (red circles), 0.24M (blue triangles), 0.34M (dark cyan diamonds), and 0.49M (pink stars).

τ_{lag}, varies with salt concentration. The results of both simulation and experiment (Fig. 3.19, Fig. 3.20) show that as the salt concentration increases, the lag time becomes more protein concentration-dependent. This makes sense since our simulations indicate that adding salt has the two-fold effect of increasing the elongation rate and shifting the CFC to lower values, so that at high salt concentrations τ_{lag} is dominated by the concentration-dependent process of fibril elongation rather than the concentration-independent process of fragmentation. However, a curious feature of the data is that the experimental lag times for 0.1 M NaCl scales more weakly than the experiments performed without salt; this feature is not captured by our simulations. More works needs to be done to better understand this behavior. It should be noted that the concentration range over which we explored the scaling of the lag-time may be insufficient to capture the

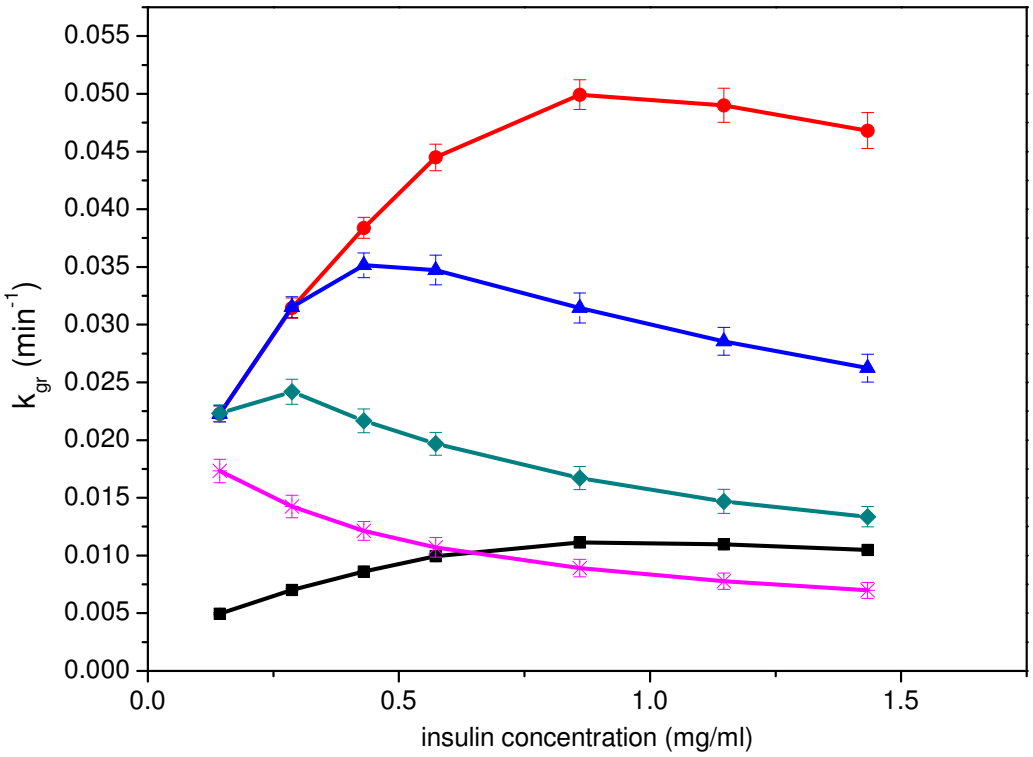

Figure 3.18: **Simulated k_{gr} obtained by varying the elongation rate and the CFC**. Parameters used in the simulations were CFC = 0.4 mg/ml, k_+=5x10$^4 M^{-1}s^{-1}$ (squares); CFC= 0.4 mg/ml, k_+=1x10$^6 M^{-1}s^{-1}$ (circles); CFC = 0.2 mg/ml, k_+=1x10$^6 M^{-1}s^{-1}$ (triangles); CFC = 0.1 mg/ml k_+=1x10$^6 M^{-1}s^{-1}$ (diamonds); and CFC = 0.05 mg/ml, k_+=1x10$^6 M^{-1}s^{-1}$ (stars). For all simulations k_f was 1x10$^{-8}s^{-1}$.

entire kinetic picture. It is not impossible that the observed scaling exponents are merely 'apparent' exponents. As has been shown clearly, fragmentation is still an active feature in the kinetics and therefore a scaling exponent of -0.5 may be recovered if a larger concentration range was explored.

3.2.6.1 Adding NaCl During the Time Course of Self-Assembly

If our model correctly represents the molecular mechanism of fibril assembly then it should be able to predict what happens when the conditions are changed *during* the process of fibril assembly. In the absence of a structural transition, one would expect that adding salt at any time before the point of maximal growth is reached will have a dramatic effect on the kinetics of fibril assembly. This is

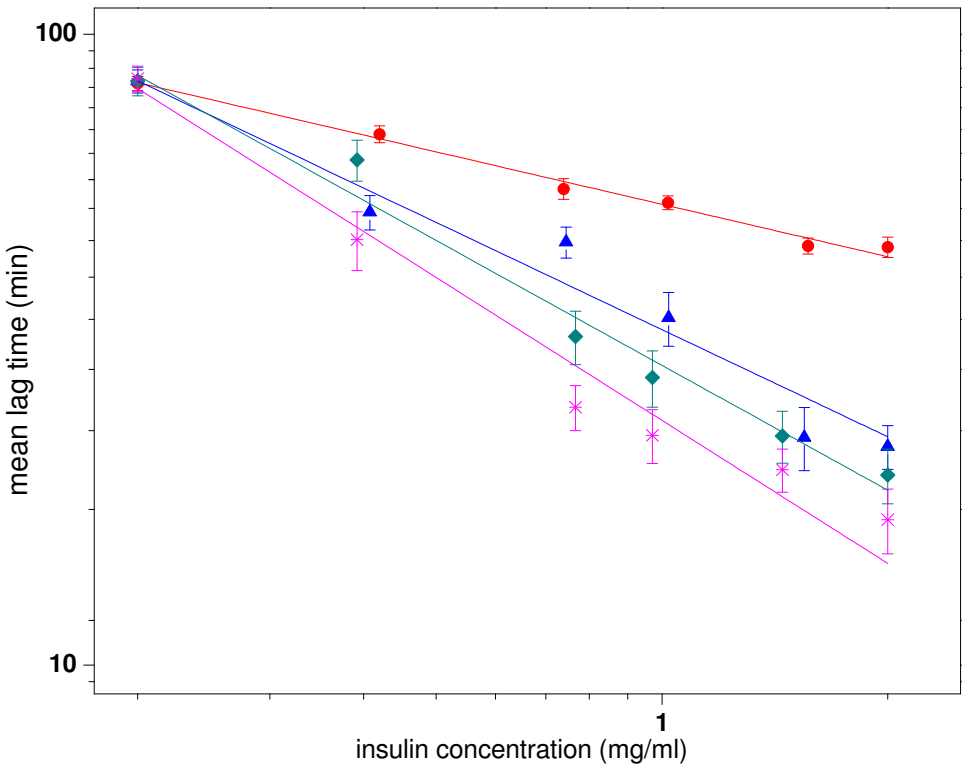

Figure 3.19: **Lag times as a function of insulin concentration in the presence of NaCl**. The data are fit same generic power law function as above. The resulting scaling exponents are -0.27 ± 0.01 (0.1M NaCl, circles), -0.56 ± 0.05 (0.24M NaCl, triangles), -0.65 ± 0.04 (0.34M NaCl, diamonds), -0.75 ± 0.06 (0.49M NaCl, stars).

because adding salt increases the fibril elongation rate, which plays a major role in determining the growth rate in fragmentation-dominated models. Plotting the growth rate as a function of the time salt is added, one would therefore expect to obtain results that follow the dashed line in Fig. 3.21. However, in a model with a structural transition, salt should not influence the kinetics if it is added after the structural transition has happened - i.e. after the fibril mass has exceeded the CFC. One would therefore expect to obtain the red line in Fig. 3.21. Indeed, this prediction is realized in our experiments. Based on our simulation results, the predicted CFC for the 0.3M salt concentration used in these experiments is 0.1 mg/ml, and that the CFC should be reached at approximately 110 minutes. Indeed, I find that adding NaCl at any time before the predicted CFC is reached (ie before 110 minutes) results in a strong increase in k_{gr}, but after 110 minutes,

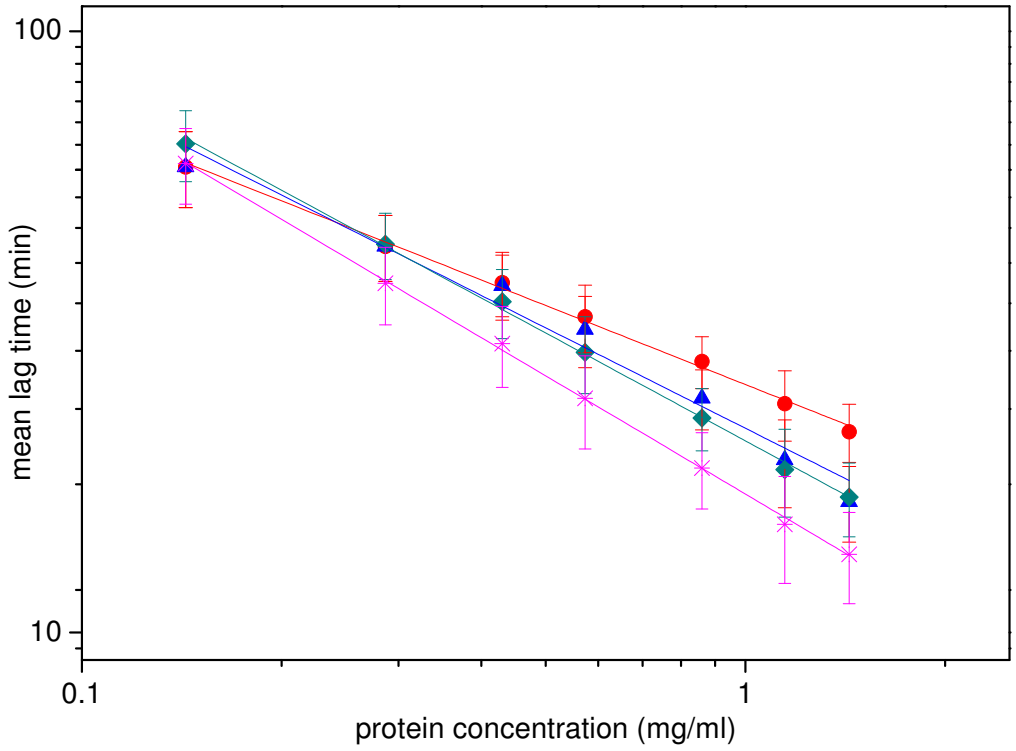

Figure 3.20: **Lag times as a function of insulin concentration obtained by varying the elongation rate and the CFC.** Parameters used in the simulations were k_+=1x10$^6 M^{-1}s^{-1}$ and k_f =1x10$^{-8}s^{-1}$. The scaling exponents are -0.43(circles) -0.55 (triangles), -0.59 (diamonds), and -0.65 (stars) when the CFC is 0.4 mg/ml, 0.2 mg/ml, 0.1 mg/ml and 0.05 mg/ml, respectively.

adding salt has no effect - i.e. our experiments indeed follow the red line rather than the dashed line in Fig. 3.21.

3.2.6.2 Implication for the existence and prevalence of small, oligomeric amyloid species

Small molecular weight species, including short fibrils, that are present during the assembly of amyloid fibrils are of particular interest because they are thought to be the causative agents of toxicity [178, 179, 180, 181]. Our model makes a strong prediction that the prevalence of short fibrils should be very different for initial protein concentrations above and below the critical threshold. Fig. 3.22 shows the prevalence of short fibril species in our simulations as a function of time. For $c_p > c_T$ we predict the transient formation of a small population of short fibrils,

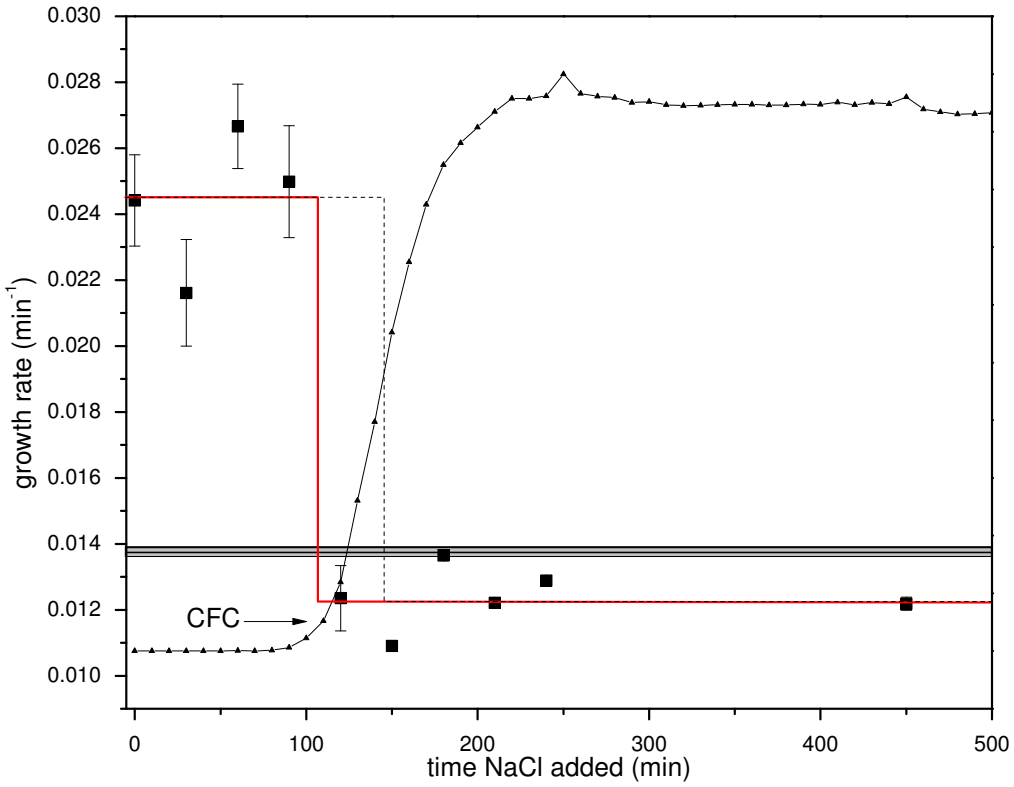

Figure 3.21: **The influence of environment on the growth kinetics.** A series of 1.5 mg/ml insulin samples were incubated simultaneously. At specific time points during the kinetics of self-assembly (every 30 minutes), NaCl was added to three samples. The mean value of k_{gr} from the resulting kinetic traces of each group of samples are plotted against the time at which the NaCl was added (squares). To illustrate where on the kinetic profile the NaCl was added, the kinetic trace for the control is also shown (triangles). The grey box corresponds to the mean $k_{gr} \pm s.d.$ for this salt-free control. In the absence of any structural transition, we would expect to observe the behavior shown by the dashed line, with enhanced growth rates during the early stages of fibril assembly, returning to control values if salt is added after the point at which k_{gr} is measured ($\sim$160 min). With the existence of a CFC, we would expect instead the trend in k_{gr} as shown by the solid line; the decrease in k_{gr} now occurs the moment the CFC is reached. Indeed, we observe that k_{gr} decreases to the control value when salt is added at any time after 120 minutes, consistent with the sample having reached the CFC, which for 1.5 mg/ml insulin under these conditions we estimate to occur at 110 minutes.

peaking approximately half-way through the exponential growth phase before returning to low levels (Fig. 3.22(B)). When $c_p < c_T$, however, the prevalence of short fibrils is much greater and increases with time (Fig. 3.22(A)). Thus the

structural transition has a dramatic influence on the prevalence of potentially cytotoxic species during amyloid fibril assembly.

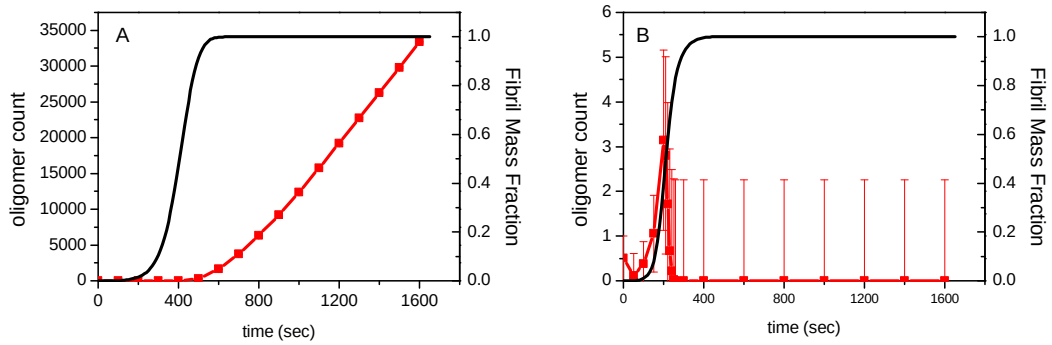

Figure 3.22: **Oligomer population as a function of time.** The number of short fibrils, of ≤ 1000 monomer units, plotted as a function of time (red squares); kinetic traces of the total fibril mass (scaled arbitrarily) are also show (black lines). (**A**) At c_p= 0.2 mg/ml, the protein concentration is below c_T and large number of short fibrillar species persist for long periods of time. (**B**) At c_p=0.55 mg/ml, a maximum in short species occurs at a time of approximately 50% assembly completion. Once the CFC is reached and the structural transition takes place, this population diminishes due to the suppression of fragmentation. The simulations are stochastic and the data points are the average oligomer number of 150 runs; the error bars are the standard deviation of the mean.

3.3 Discussion

The key finding in this study is the identification of a critical fibril mass concentration above which insulin amyloid fibrils undergo a structural transition that suppresses fragmentation. By incorporating this feature into our computational

models we are able to successfully simulate the complex behavior that we observe in our kinetic experiments under a variety of solution conditions. Indeed, the length distributions we found from simulation and TEM experiments for fibrils formed above the critical concentration are similar to those found in previous experiments [171, 172, 173] while both simulations and our TEM experiments show very different, exponential-like length distributions at protein concentrations below the predicted structural transition.

It has become the general consensus that small oligomeric aggregates are the causative agents of cytotoxicity [178, 179, 180, 181, 41] while the mature, long fibrils are the inert end product of assembly [182, 183]. Importantly, Xue and co-workers have demonstrated that it is not just the oligomers formed during self-assembly that are toxic, but that short fibrils generated through the fragmentation of mature fibrils also cause cell death [184]. Intriguingly, multiple studies have shown that there is no direct correlation between fibril load within tissue (i.e. mass of long fibrillar material) and loss of function ([185, 186, 187, 188, 189, 190, 182, 183]) Indeed, it has been shown that a significant fraction of members of the Nun study [191] possessed neuritic β-amyloid plaques and neurofibrillary tangles at the time of death without exhibiting any cognitive dysfunction. The observation that in some cases fibrils can be present but not cause measurable functional impairment could be due to amyloid fibril assembly pathways that favor the rapid formation of longer inert fibrils as a result of the local intracellular environment. Our observation of two distinct length distributions that are determined only by the environmental conditions, one of which is comprised of long, potentially inert fibrils microns in length while the other is dominated by very short, possibly toxic, fibrillar species, may have consequences for determining the propensity of intracellular conditions to bias toward either benign or cytotoxic fibrillar loads.

To this end, we have shown that salt concentration has a strong impact on the value of the critical fibril mass concentration. Higher salt concentrations shift the CFC to lower values, stabilizing the length distribution for lower initial concentrations of protein. In our study we chose salt as a tool to alter protein-protein and fibril-fibril interactions, but any additive or solution condition that can affect the charge characteristics or protein structure, such as pH, temperature, or the presence of hydrophobic surfaces, could affect the value of CFC. This could

have far-reaching consequences for toxicity in intracellular environments, where salt gradients, pH differences, the presence of chaperones and other co-factors could bias toward the formation of toxic or non-toxic fibrillar species.

It should be noted that the trend we nominally observe of k_{gr} as a function of concentration has been observed for a number of polymerizing fibrillar proteins, in particular insulin, bovine lactalbumin [192], and yeast prion protein [76]. In these cases, above a certain concentration of protein, the scaling of the growth rate becomes effectively zeroth-order. Formally, such behavior can be understood as an analogue of enzyme reaction kinetics described by a Michaelis-Menten type model. In this case, the substrate is the reactive monomer species and the enzyme is the fibril end. A concentration independent reorganization step occurs after monomer binding to the fibril end resulting in the apparent concentration independence in the observed growth rate. This observation does not negate the interpretation of the data as presented but serves as a complement to the phenomena that is described in this chapter. Indeed, the fact that k_{gr} seems to *decrease* with increasing insulin concentration is consonant with our model and could not be captured by a Michaelis-Menten type model.

Interestingly, Taylor *et al.* map the level of Aβ(1-40) cytotoxicity to the kinetic profile of fibril formation [9], reporting a peak in cytotoxicity during the exponential growth phase of assembly (Fig. 3.23). After this peak is reached, cytotoxicity decays back to baseline levels at the point in the kinetic profile where fibril assembly approaches and reaches the final plateau. The results of Taylor *et al.* may be interpreted in the context of our structural transition model: a population of small molecular weight species is generated through fragmentation, and then depleted by the combined influence of fragmentation inhibition due to a structural transition and fibril elongation. Applying our model to the data presented in [9] we would predict the CFC for Aβ(1-40) to be approximately 0.15 mg/ml($\sim 35\mu$M) under these experimental conditions. In our simulations at high protein concentrations, the prevalence of short, potentially cytotoxic oligomeric species shows very similar kinetics to the kinetics of cytotoxicity reported by Taylor *et al* (Fig. 3.22(A)). For simulations in which the initial protein concentration is above the CFC we find an analogous peak in small molecular weight species mid-way through the exponential growth phase which is followed by a decay back to baseline levels. In contrast, when the concentration

of protein is below the c_T, the residence time of small, potentially toxic species is significantly longer: indeed they remain present in high abundance even at very long times (Fig. 3.22(B)). This could have profound consequences for the propensity for cellular damage or death when intracellular conditions are such that short fibrils are the more likely species to be present. If this is indeed the case, then it may be possible to devise therapeutic measures that take advantage of these insights to alter cellular conditions so as to bias toward the accumulation of longer inert fibrils and away from the shorter toxic species.

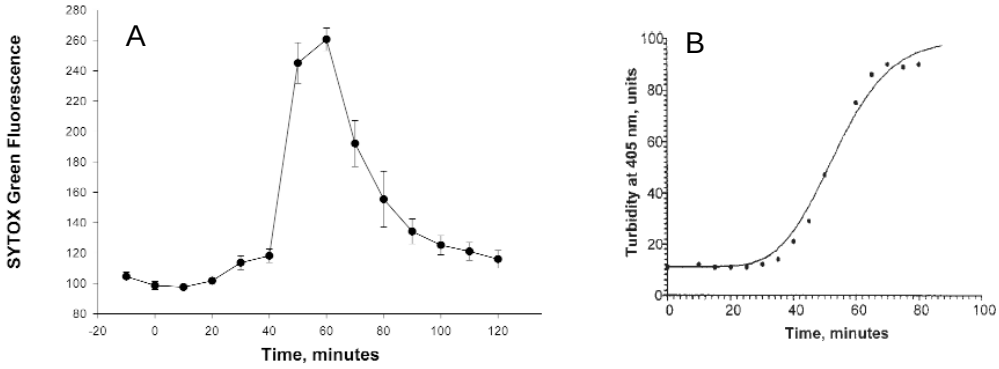

Figure 3.23: **A measure of cytotoxicity as a function of time during Aβ(1-40) fibril formation.** (**A**)Taylor *et al.* quantified cytotoxicity by measuring SYTOX green uptake by human epithelial cells (SH-EP1). SYTOX green is only taken into the cell when the cell membrane becomes permeable. Upon uptake, SYTOX green binds to nucleic acids and has a marked increase in fluorescence intensity. (**A**) shows the level of SYTOX green fluorescence as a function of Aβ(1-40) fibrillation time. (**B**) is a plot Aβ(1-40) fibril formation kinetics as monitored by turbidity at 405 nm. One observes that a peak in cytoxicity in (**A**) occurs at approximately the point of maximum aggregation in (**B**). Taken from [9].

Amyloid and amyloid- like fibrils have also received attention for their potential use as biomaterials or components in composite materials (see §1.3.5, [163]). For example, Reches and Gazit have used fibrils to template metal nanowires [77], Channon *et al.*, has used them as the basis for light harvesting structures [86], and a number of authors (e.g. Zhang [75, 161] and Gras *et al.* [66]) have used them as the basis for cell scaffolds. A significant challenge remains in controlling the length distribution of fibrils, and the viscoelasticity of the resulting hydrogel. Our current findings suggest mechanisms for manipulating both, and

for predicting and controlling the homogeneity of the resulting population.

The exact nature of the structural transition has yet to be determined. Certainly, the CD spectral data is indicative that there is potentially a complex conversion of fibril architecture taking place. Chapter 4 will begin to attempt to address this issue by studying fibril formation using small angle neutron scattering. It remains an experimental and theoretical challenge to understand the fundamental origin for the observed transition, and the subsequent arrest of fragmentation, that we see in our experiments.

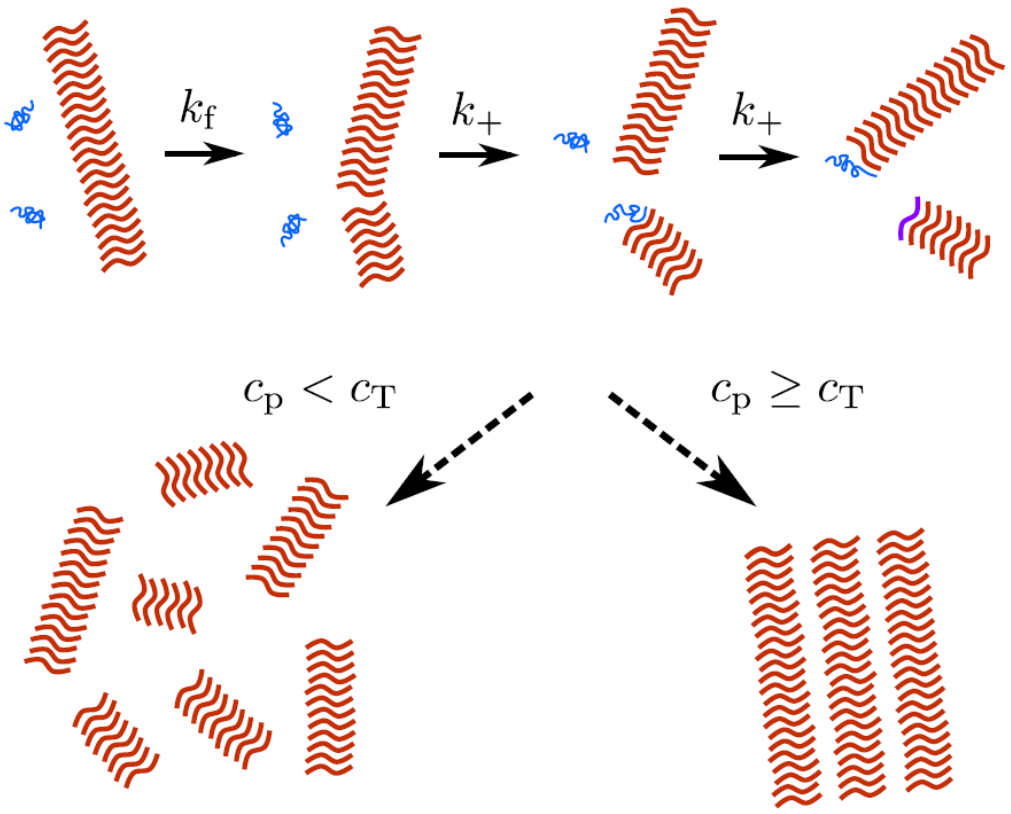

Figure 3.24: **Schematic of amyloid fibril assembly with the presence of a CFC.** Amyloid fibrils grow by monomer addition to the ends of the fibrils. Fibril fragmentation multiplies the number of free ends by which monomer may attach. For concentration below the the critical threshold concentration, c_T, fragmentation is never arrested. The result are short fibrils at long times. In contrast, when the concentration is above c_T, fibril-fibril association arrests fragmentation resulting in a broader distribution of long fibrils.

3.4 Conclusions

We have shown evidence for a structural transition which occurs at a critical fibril mass concentration, or 'CFC', above which fragmentation of the fibrils is suppressed. Fig. 3.24 is a schematic of the features of this model. Our simulations predict the formation of distinct fibril length distributions for protein concentrations above and below a critical threshold, which we confirm by electron microscopy. These results point to a new picture of amyloid fibril growth in which structural transitions that occur during self-assembly have strong effects on the final population of aggregate species with small, and potentially cytotoxic, oligomers dominating for long periods of time at protein concentrations below the critical threshold, but not for protein concentrations above the threshold. We further show that the CFC can be modulated by environmental conditions, pointing to possible *in vivo* strategies for controlling cytotoxicity.

Chapter 4

Neutron Scattering Studies of Insulin Amyloid Fibril Formation

4.1 Introduction

In Chapter 3 we observed the existence of a structural transition where the physical mechanisms that drive amyloid fibril assembly are altered due to a physical arrest of the system. The exact mechanism of why fragmentation becomes suppressed has yet to be determined. It could be due to fibril thickening and/or bundling via fibril-to-fibril interactions [137], the formation of a gel-like entangled network of fibrils [174], the association of fibrils into larger macro-fibrillar assemblies such as spherulites [193, 194], or a combination of these processes. The ultimate goal is to understand, on a mechanistic level, this structural transition. Small angle neutron scattering (SANS) is well suited to address these questions because it provides a means to study *in situ* the evolution of structural changes in the aggregates in a time-resolved manner over a range of length scales from angstroms to hundreds of nanometers. A secondary advantage is that SANS allows one to carry out time-resolved biological experiments without fear of significant radiation damage to the sample (i.e. as can be a problem with SAXS). SANS has been used to study the evolution of aggregates of insulin amyloid fibril assembly [195, 196] and other protein amyloid systems [111, 197, 198]. In this Chapter I present initial results obtained from experiments performed on the Sans2D instrument at ISIS, Rutherford-Appleton Laboratory (see §2.2.7.2 for details on this instrument). We measured the aggregation of

insulin into amyloid fibrils, in a time-resolved manner, over a wide Q-range (0.005-0.8 $Å^{-1}$) so that the structural evolution of the aggregates could be monitored and characterized over multiple length scales in order to shed light on the mechanisms of the structural transition we observed in Chapter 3.

4.2 Results

4.2.1 Time-resolved SANS of 4 mg/ml insulin amyloid fibril samples

We can model the expected neutron scattering signal for insulin under our experimental conditions (see §2.2.8.3). We expect insulin to predominately be monomeric [39, 199]. So that a theoretical scattering curve can be obtained, we can geometrically approximate the insulin monomer can be modeled as a cylinder with a length L=30 Åand radius r=11 Å(the dimensions of which are obtained from crystal structure of an insulin monomer (PDB 1GUJ), Fig. 4.1). The results are shown in Fig. 4.2. Using these dimensions one can obtain the radius of gyration (R_g) for an insulin monomer, given by the expression,

$$R_g = \sqrt{\frac{r^2}{2} + \frac{L^2}{12}} \tag{4.1}$$

which yields a value of (11.6 Å). This result is in good agreement with both experiment values (11.9 Å[39]) and simulation (10-13Å) [200, 201]. While it is a fair approximation to model the system purely as monomers, it must be acknowledged that in reality a solution of insulin is not entirely monodisperse. It has been shown that even at $t = 0$, there exists a distribution of not only oligomers but also multiple conformations of each species [201]. With this in mind we may attempt to fit the experimental scattering curves at t=0 to this approximate cylindrical model (§2.2.7.3) Fig. 4.3 is the neutron scattering signal for a 4 mg/ml insulin sample at the beginning of a kinetic run. We find a close agreement between the cylindrical model and the observed signal. Fitting this signal with a cylindrical model yields $R_g = 11.67 \pm 0.98$Å. Despite the polydispersity at early stages of the kinetics, this simple cylindrical model of the monomer fits well to the data.

We monitored the kinetics of fibril assembly over the course of several hours.

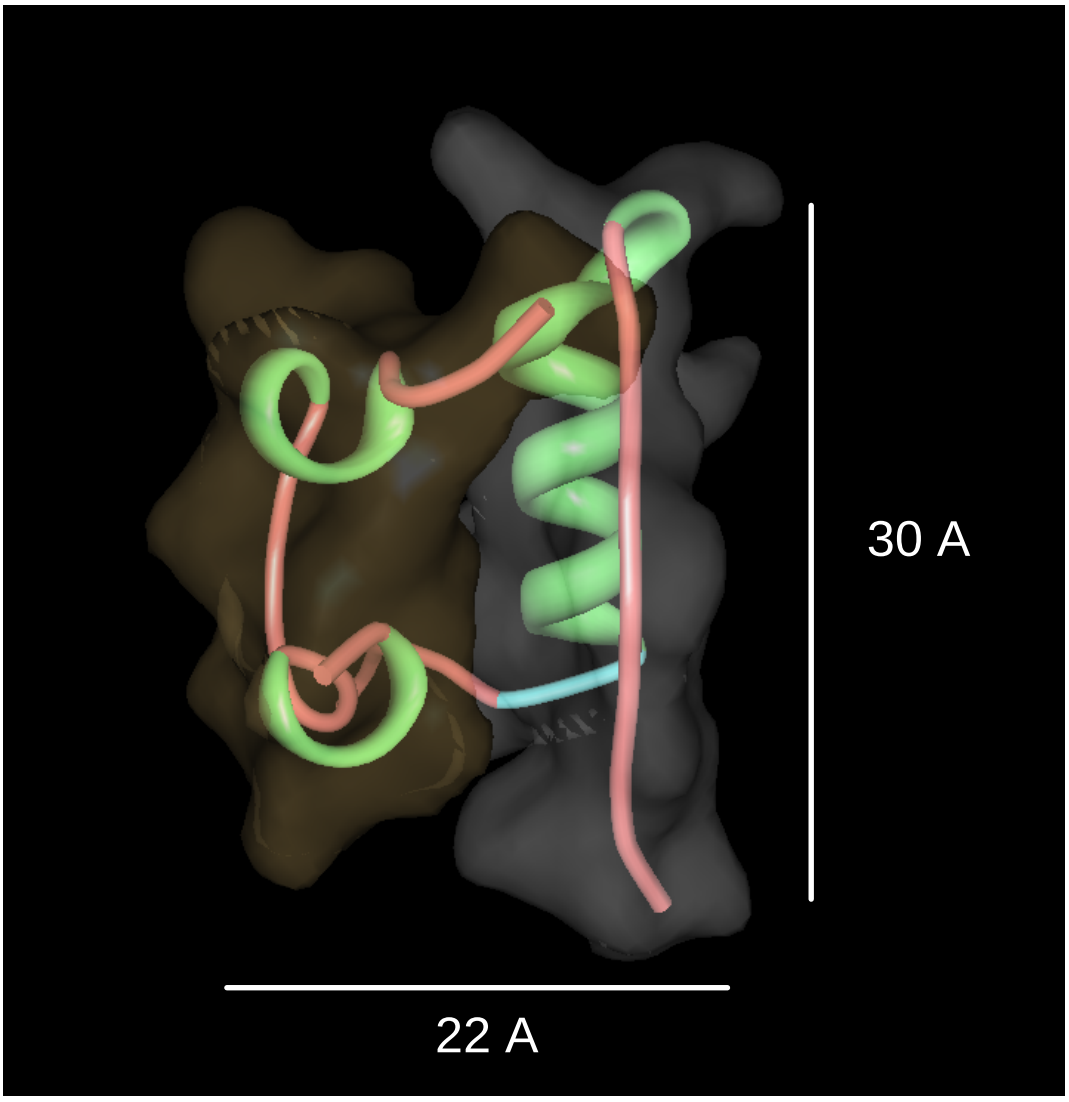

Figure 4.1: **Dimensions of the insulin monomer.** Shown is a cartoon of the insulin monomer taken from PDB 1GUJ. The dimensions used for the theoretical R_g are r=11 Åand L=30 Å. The helices are green, coil are red, turns are blue. Surrounding the cartoon is the surface representation of the protein where yellow is the A-chain and white is the B-chain.

The resulting neutron scattering profiles are shown in Fig. 4.5. For the first $\sim$ 150 minutes we find that the scattering profile does not change. For t $>$ 200 minutes we observed increasing scattering intensity at lower Q-values; this trend increased with time (see Fig. 4.5). We first attempted to fit the data with the cylindrical model (examples for t=0 and t=225 mins are shown in Fig. 4.8 A and

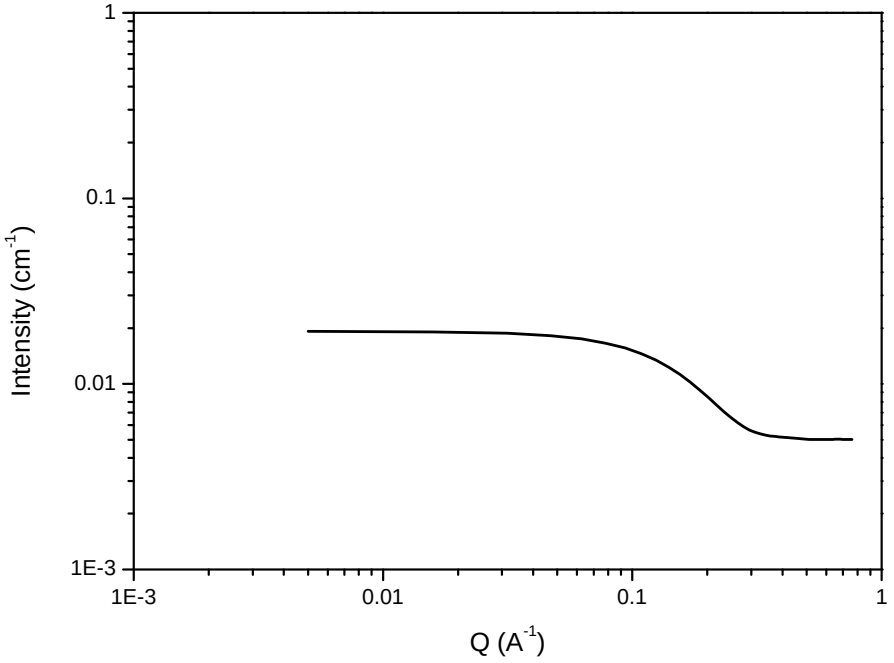

Figure 4.2: **Scattering function for a cylinder with a dimensions L=30År=11Å** In order to obtain a theoretical scattering curve we assume the shape of the monomeric state of insulin to be approximated by a cylinder. We use the crystal structure dimensions of an insulin monomer (PDB 1GUJ) as inputs for the simulation. The results are shown in the figure above for the Q-range explored in experiment.

B). For times $0 \leq t \leq 255$ mins, the data could successfully fit to the cylinder model. The fits to the data for these time points give lengths that are constant until t=180, at which point there is an increase in the length of the cylinders (Fig. 4.6). The radius of the cylinders remain constant over this time interval. However, for $t \geq 270$ mins, the cylindrical model fails to fit the data, particularly at higher Q-values (Q > 0.1 Å^{-1}). This is demonstrated in Fig. 4.4, where the cylinder model is fit to the data obtained at $t = 270$ min. One can see that the model does not fit the data well at higher Q-values (χ^2/N =3.2). This feature persists for scattering data obtained at times greater than 270 min.

In an effort to understand what does explain the data we plotted the real-space probability distribution of scattering elements $P(r)$. The distribution $P(r)$ can provide information about the amount of polydispersity in a system. Fig.

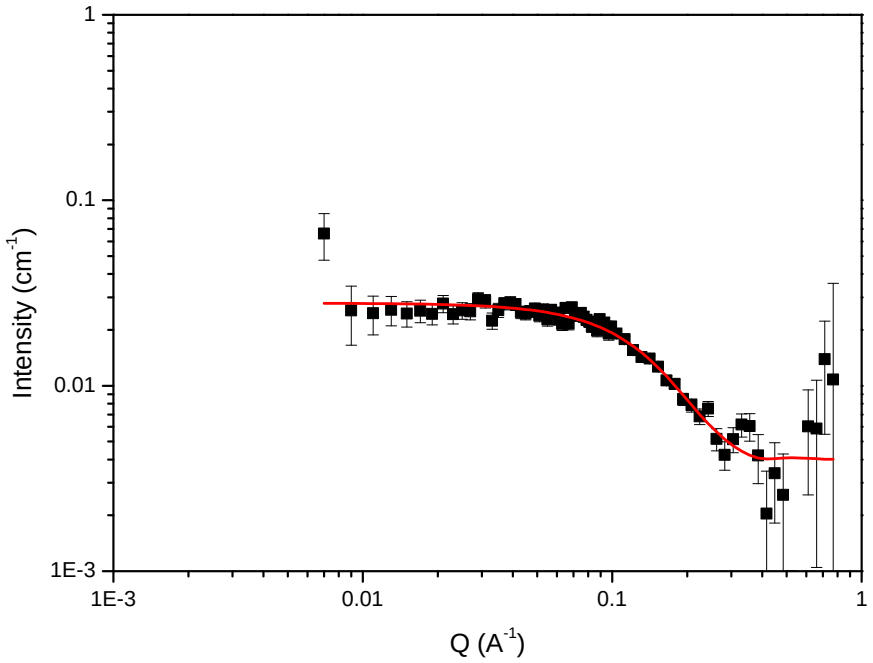

Figure 4.3: **Scattering function of 4 mg/ml insulin at $t = 0$.** Plotted is the scattering curve for 4 mg/ml insulin at at the beginning of the kinetic run. We find that this scattering curve is well-fitted to a model where the scatterer is cylindrical in shape. The red-line is the line of best fit which yields R_g=11.67±0.98 Å.

4.7 shows $P(r)$for several time points in the kinetics. It is apparent that at early times, the system is quite monodisperse and gives a fitted R_g value of ∼ 12 Å. Moreover, the fitted value for I(0)=0.023 $Å^{-1}$ yields a molecular weight M_w=5678±50 Da, which is within error of the M_w for monomeric insulin (5733.5 Da), where the protein specific volume, ν= 0.73 cm^3g^{-1} and the contrast of scattering length density between the protein and the solvent, $\Delta\rho$=3.4$x10^{10}cm^{-2}$ (see §2.2.7.3). At t=180 and t=225 min, there are additional peaks at larger r-values indicative of aggregate species present within the sample. However, despite the development of increasing polydispersity by the formation of larger aggregates within the system, the cylindrical model still fits reasonably well to the data at these times, perhaps due to the still extant monomeric peak. We find that the time at which the cylindrical model fails is when the monomer peak effectively disappears and is overwhelmed by larger aggregate species. By the

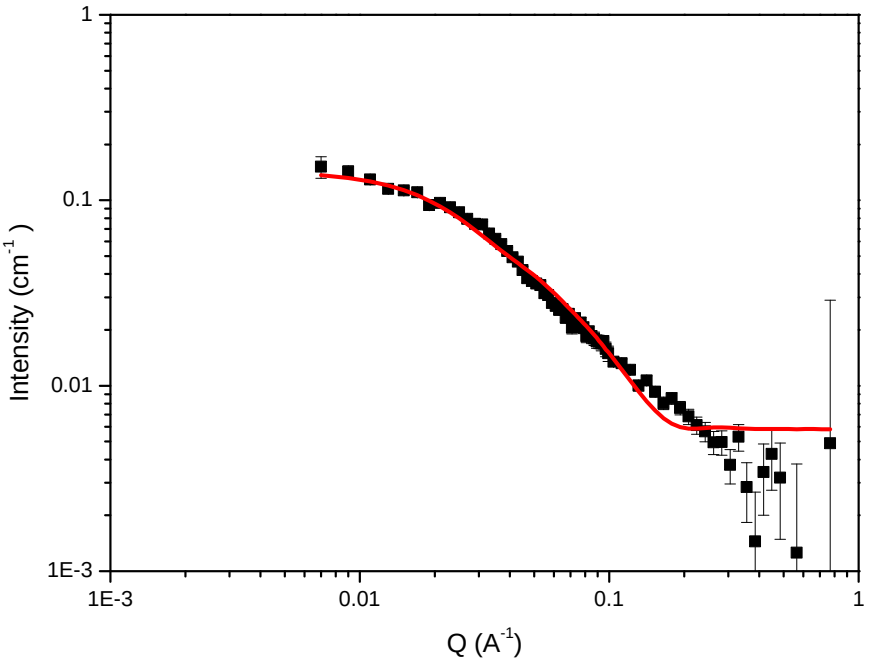

Figure 4.4: **Fit to the cylinder model at t=270 min.** Plotted is the scattering data at 270 min fit with the cylinder model. Clearly, the fit is poor, especially at high Q-values. This lack of fitting to the cylinder model persists for times greater than 270 min.

final time point, a very broad distribution of large aggregate species dominates the scattering signal.

Clearly, it is no surprise that for t $\geq$ 270 mins, the data can not be modeled by a simple model. In an effort to extract some quantitative sense of the structure of the aggregates we approximate the scattering signal by a combination of more than one cylindrical species. The justification for this is that the signal for t $\geq$ 270 mins is still growing therefore monomers and their signal must still be present. This proposition is plausible since many models of insulin fibril assembly involves elongation via monomer addition [202, 119](§1.3 & §1.4). Obviously, the fit to larger objects will be effected by polydispersity. Nevertheless, we fit the data to a simple linear combination of a cylinder and flexible cylinder (§2.2.7.3) and examples are shown in Fig. 4.8C and D. The results of the fit gives two lengths L_1 and L_2 and two radii, r_1 and r_2. For t=270 mins, the large Q-values are fit much

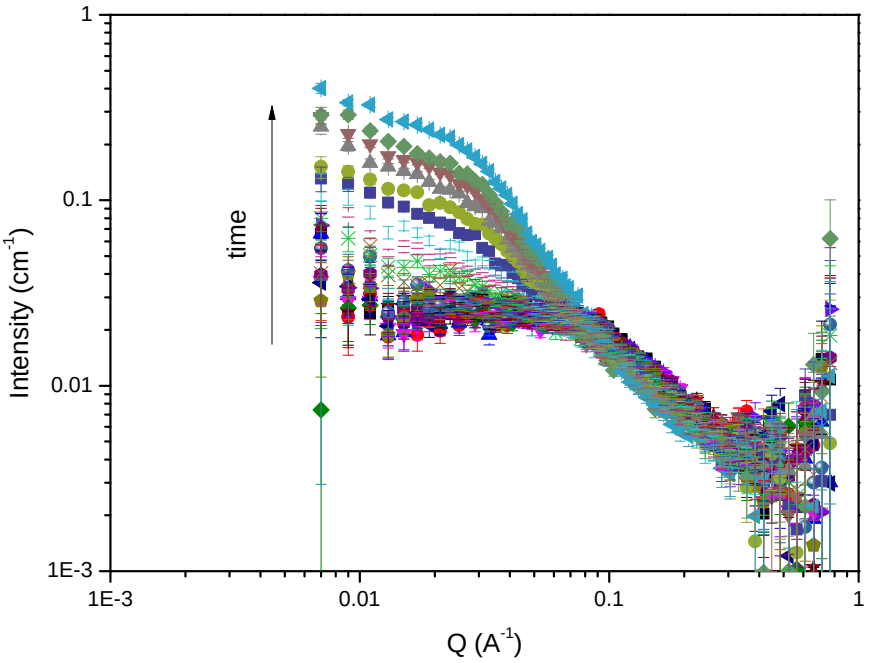

Figure 4.5: **Time-resolved scattering curves for 4 mg/ml insulin.** Plotted are the scattering curves for 4 mg/ml insulin over a time range 0 to 345 minutes. For t<150 minutes the scattering curves remain constant. For t>200 minutes an increase in the scattering intensity is observed in the low-Q range of our experiments corresponding to growth of larger aggregates.

better by this model yielding $L_1 = 981 \pm 132$Å, $r_1 = 35 \pm 12$Åand $L_2 = 36 \pm 5$Å, $r_2 = 9 \pm 3$Åwith a χ^2/N=0.77. At the final time point t=345 mins the fit can be seen to not be as good with a χ^2/N=1.77. However, the values obtained are $L_1 = 2484 \pm 345$Å, $r_1 = 36 \pm 15$Åand $L_2 = 51 \pm 12$Å, $r_2 = 15 \pm 6$Å. For these time points the aggregates appear to become thicker, perhaps due to fibril-fibril interactions. It should be noted that lengths beyond these will begin to move outside the Q-window of these experiments; larger species will no longer be able to be distinguished.

One must make mention that the time scale of fibril assembly in these *in situ* experiments are not the same as in Chapter 3. Such differences could be associated to inefficient heating of the sample within the chamber. Another source of discrepancy could be due to cuvette geometry and the type of material it is

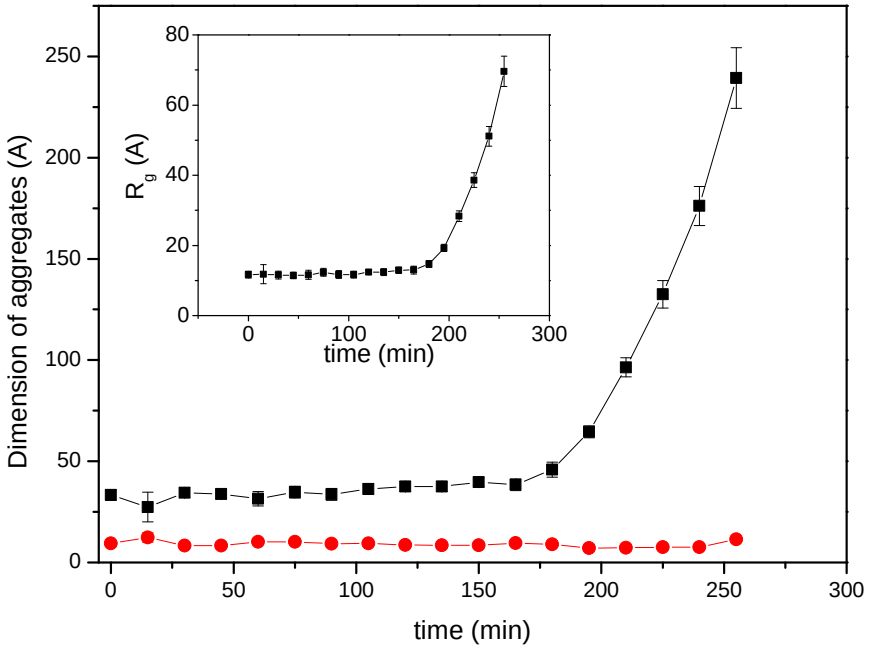

Figure 4.6: **The length, radius, and R_g as a function of time for times $0 \leq t \leq 255$ mins.** Plotted are the lengths (black squares) and radii (red circles) obtained from fitting the data up to 255 min by a cylindrical model. The fitted lengths are constant until t $\geq$ 180 mins, after which there is an exponential increase. The radii over the same time remain constant. The inset shows R_g for a cylinder as a function of time obtained from the fitted values of L and r.

made of; as was discussed in §1.2, the existence of hydrophobic surfaces can influence the kinetics of aggregation.

Regardless of the efficacy of fitting a model to the data, what is incontrovertible is that at a critical time (t$\geq$ 270 mins) in the aggregation kinetics, there is a marked increase in the scattering at low Q-values due to the formation of larger aggregates. It would be extremely interesting to compare the time at which this happens to the percentage of aggregate formed. According to the work in Chapter 3, the CFC for insulin under these experimental conditions is 0.4 mg/ml which corresponds to 10% of monomer being converted to aggregated material. However, it is difficult to determine how much material has become fibrillar because the experiment unfortunately was interrupted by a major fault in the LINAC. It is therefore difficult to get an estimate of the percentage of

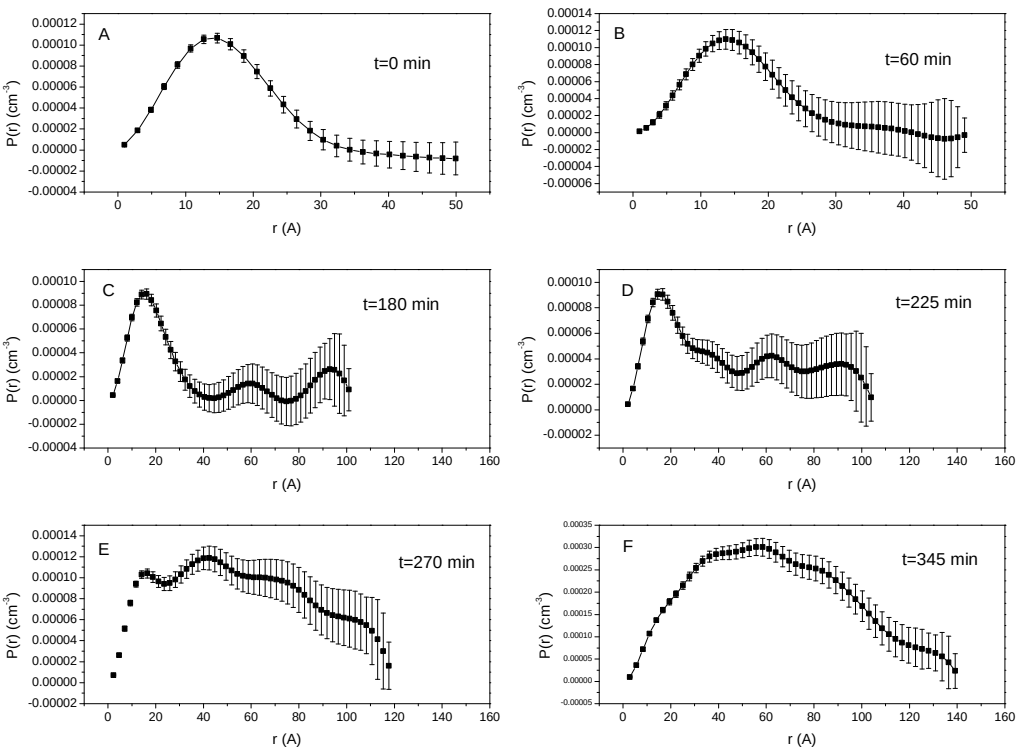

Figure 4.7: $P(r)$ **distributions for several time points in the kinetics.** Plotted are real-space distributions $P(r)$ for (**A**) 0 mins, (**B**) 60 mins, (**C**) 180 mins, (**D**) 225 mins, (**E**) 270 mins, and (**F**) 345 mins. Note the change in x-axis between 60 and 180 min. At t=0, the solution is dominated by a population of monomer-sized scatterers. Over time, larger-sized aggregates form until a critical time at t=270 mins (which corresponds to a simple cylinder model no longer fitting the data) when the monomer peak is dominated by larger aggregates. By the final time point at 345 mins a large distribution of large aggregates are present and the monomer peak is obscured completely.

fibrillar material in the system at the critical time. However, with a complete kinetic run the determination of fibril mass percentage should be achievable.

4.2.2 SANS of 0.5 mg/ml insulin amyloid fibrils

We also attempted to look at the scattering of a 0.5 mg/ml insulin sample. This concentration was chosen because it is below the critical concentration at which we expect a structural transition from to occur (§3.2.1.3). The results are shown in Fig. 4.9. We find that, as expected, the signal-to-noise ratio is much less favourable at this lower concentration. Despite this, a scattering signal is

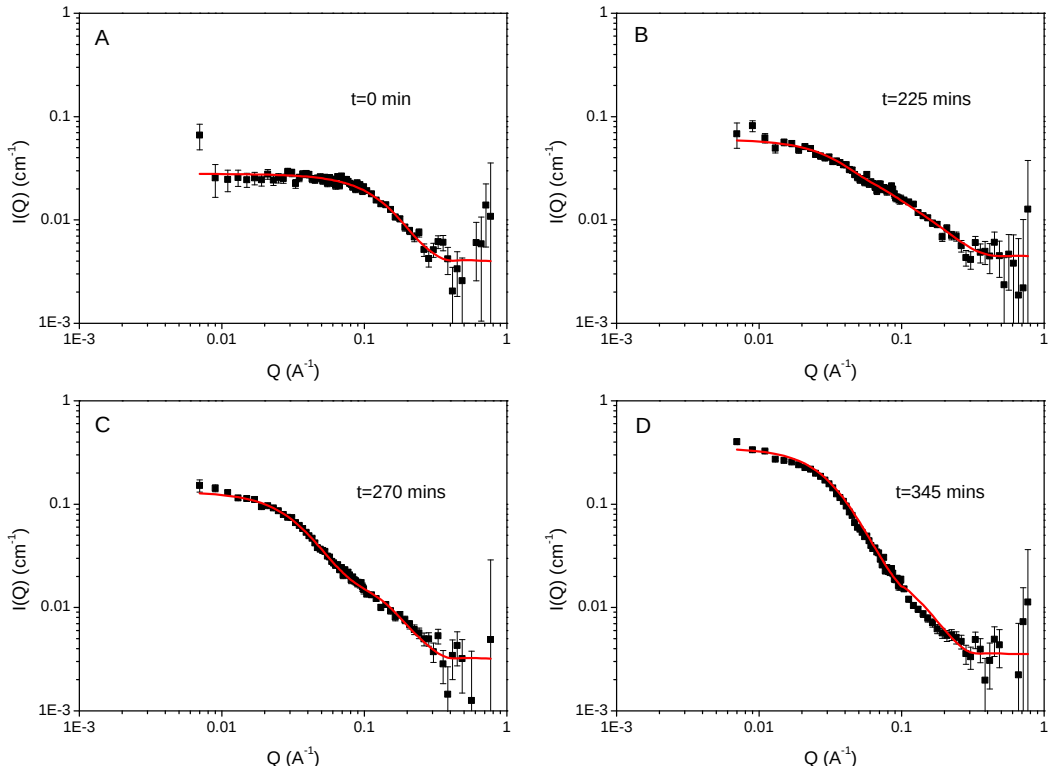

Figure 4.8: **Examples of the data fit to cylinder or cylinder+cylinder models.** (**A**) t=0 and (**B**) t=225 mins scattering curves fit to a single cylinder model. (**C**) t=270 and (**D**) t=345 mins scattering intensities are fit to a linear combination of two cylinder models.

discernible. We allowed the sample to incubate for over 15 hours. Over that time we did not observe any evolution in the scattering signal.

Even though we did not observe an evolution of the scattering signal for this lower concentration we did look at the scattering from a sample that had been prepared and pre-incubated in a 96-well plate. We prepared two sets of samples: 1) 0.5 mg/ml insulin in standard buffer and 2) 0.5 mg/ml in buffer with the addition of 100mM NaCl. Salt was added to the sample as it is expected to shift the CFC to a lower value (§3.2.6). The sample was incubated for 8 hours at 60°C. The resultant scattering profiles are taken from samples that have reached the plateau region (§3.2.1). The results are shown in Figs. 4.10 and 4.11. We fit the resulting scattering signals with a power law of the form $I(Q) = AQ^{-m}$ where Åis a scale factor and m is the scaling exponent. The data was fit from 0.05 to 0.5 $Å^{-1}$ (one decade). The scaling exponent m is related to the fractal dimension

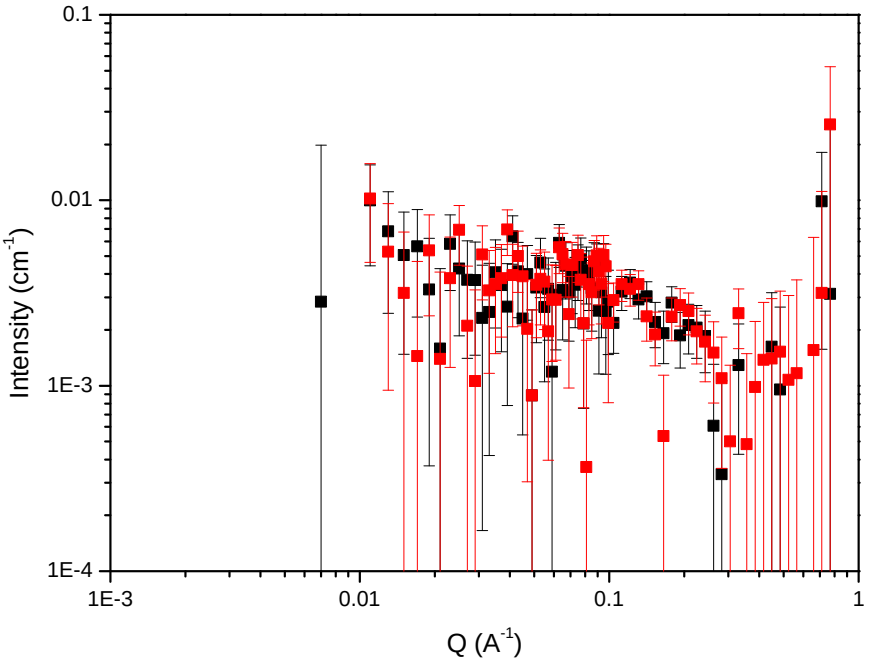

Figure 4.9: **Time-resolved scattering curves for 0.5 mg/ml insulin.** Plotted are the scattering for 0.5 mg/ml insulin at t=15 minutes and t=15 hours, 10 minutes. The signal-to-noise ratio is much less favourable due to the sample being much more dilute. We still find a scattering curve that is discernible over the background. However, in the time-scale of these experiments we did not observe any evolution in the scattering function.

of the scattering object (see §2.2.7.1, [203]). We find the that $m = 1.66 \pm 0.04$ for the sample without NaCl, and $m = 2.12 \pm 0.07$.

Interestingly, the scaling exponent for the sample without NaCl has an exponent in good agreement with the scaling exponent that corresponds to a self-avoiding polymer chain ($m = 5/3$). Scaling exponents in the range $2 \leq$ m ≤ 3 are typically indicative of mass fractal structures such as branched systems or networks, i.e. gels. This observation fits well with what would be expected based upon the model outlined in Chapter 3. Specifically, in the absence of NaCl, the critical concentration, $c_T \sim 0.75$ mg/ml (§3.2.1.3). It was shown in §3.2.3 that, an initial concentration of 0.5 mg/ml will not result in the formation of a gel/viscous fluid. However, with the addition of NaCl, it is expected that the critical concentration, c_T will shift to lower values (§3.2.6). Quantitatively, the

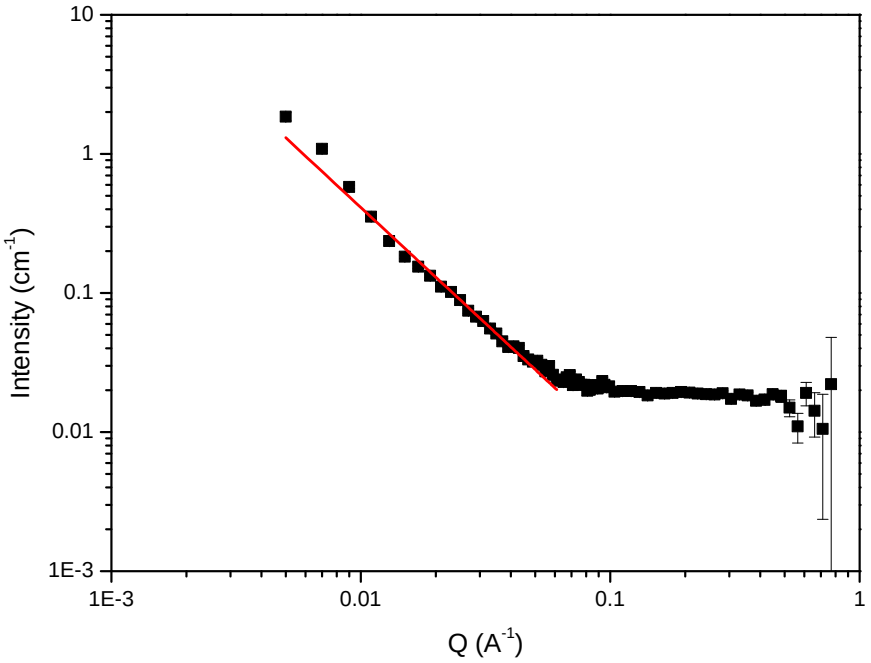

Figure 4.10: **Scattering intensity for the end-point of 0.5 mg/ml fibril kinetics.** Plotted is the scattering intensity for 0.5 mg/ml insulin sample that had been incubated in a 96-well plate at 60°C for 24 hours. The data is fit to a power-law with an exponent of -1.66±0.04. Fit gives an adjusted R^2=0.96.

results in that section showed that for a concentration of 0.1M NaCl, $c_T \sim 0.4$ mg/ml. Therefore, it would be expected that for initial insulin concentration above this value, a gel-like state should be observed. Indeed, the scaling behavior of the 0.5 mg/ml sample in the presence of 0.1M NaCl is 'gel-like' and is in accord with this prediction.

It should be noted that the scattering intensities at high Q-values for these incubated samples (Figs. 4.10 & 4.11) is higher than data collected from samples formed *in situ* (Fig. 4.5). This is most likely due to exchange of D_2O buffer with H_2O from the atmosphere so that simple background subtraction of D_2O buffer does not completely remove higher incoherent background scattering from H_2O. This can be resolved in the future by taking background measurements of mixtures of different percentages of D_2O/H_2O buffer.

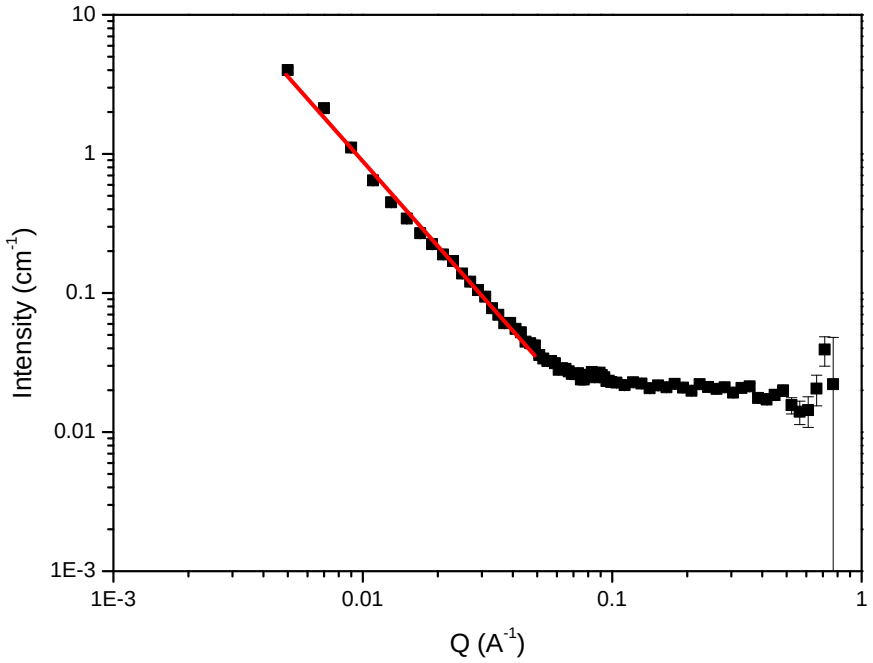

Figure 4.11: **Scattering intensity for the end-point of 0.5 mg/ml, 0.1 M NaCl fibril kinetics.** Plotted is the scattering intensity for 0.5 mg/ml insulin sample with the addition of 0.1 M NaCl that had been incubated in a 96-well plate at 60°C for 24 hours. The data is fit to a power-law with an exponent of -2.19 ± 0.07. Fit gives an adjusted R^2=0.98.

4.3 Discussion & Conclusion

We have used time-resolved SANS using the Sans2D detector at the ISIS facility to monitor the aggregation of insulin into amyloid fibrils. We observed, for an insulin concentration of 4 mg/ml, an initial scattering curve that was indicative of a population of insulin monomers with a R_g of $\sim$ 11 Å. The monomeric insulin scattering curve was well described by assuming cylinders as the scattering centres. Over time, we observed increased scattering intensity at low-Q values and decreased scattering at high-Q. We found that at a critical time, the data was no longer described by a single cylinder model. This time corresponds to a marked increase in polydispersity and the loss of a discernible monomer signal. We fit the data for times greater than 270 mins with a modified model which is simply a linear combination of cylinder models. The fits indicate these aggregates

have larger diameters when t $\geq$ 270 mins relative to the aggregates formed when t$<$ 270 mins.

During the acquisition of this data at ISIS there was a major vacuum fault in the linear accelerator which prevented further data from being acquired. Nevertheless, the preliminary data here demonstrates that insights may be gained concerning the structural evolution of insulin amyloid aggregates using SANS. The primary objective is to understand the structural transition that we describe in Chapter 3. It is of great interest to see whether the increase in scattering observed at 270 mins corresponds to the time when the CFC is reached. In order to determine accurately the percentage of fibrillar aggregate present, a full kinetic SANS run needs to be obtained. The subsequent full kinetics could be measured either *in situ* or by taking samples that are incubated in 96-well plates. This latter approach may be more advantageous since it would correspond more directly to the work done in Chapter 3. One obvious disadvantage of this approach would be the possibility shear-induced alterations of the fibrils due to pipetting the samples from plate to cuvette. What this work has demonstrated is that SANS can effectively monitor structural changes in aggregating protein samples. With complete kinetic SANS data one may be able to map the alterations in structure to the model outlined in Chapter 3.

Of course it would be useful to obtain cleaner scattering data for low insulin concentrations (particularly below the c_T). This will require longer acquisition times, however, with a cost of decreased kinetic resolution. Alternatively, the use of thicker sample cells may help ameliorate the issue. If this is achievable it would be interesting to measure the mass per unit length, as a function of time, for aggregates above and below c_T. In light of the CD results in Fig. 3.10, where we speculate that the shift in the minima at low concentrations is due to fibril-fibril association, such analysis could go toward confirming this hypothesis. Nevertheless, it is extremely encouraging that the scaling behavior of the 0.5 mg/ml samples with and without NaCl are concordant with what would be expected from the CFC model discussed in Chapter 3.

Chapter 5

Investigation of the Kinetics of Loop Forming Fibrils Using Ovalbumin

5.1 Introduction

Amyloid fibrils display a large diversity of morphologies depending on the protein or peptide of interest or due to the conditions in which they form [53, 58, 52, 107, 137]. It was observed that highly flexible amyloid fibrils formed from Apolipoprotein C-II (apoC-II) assemble into 'circularized' closed-loop amyloid fibrils [56, 57]. In this chapter, I explore the kinetics of loop-forming fibrils utilizing ovalbumin as a test protein system. Under the correct conditions ovalbumin forms closed-loop fibrils. I find interesting characteristics in the kinetics of ovalbumin fibrils formed under our experimental conditions. The kinetics are typified by the existence of three features: the lack of a lag time; immediate onset of rapid exponential growth; and very slow growth at long times. These features in the kinetics are in contrast to the kinetics of insulin fibril formation that was discussed in Chapter 3: a lag time followed by exponential growth phase which completes in a final 'plateau' in the growth profile. I used ThT fluorescence assays to monitor the growth kinetics of ovalbumin fibrils and TEM to observe their subsequent morphological characteristics. I write down a simple model that, by including a term which takes into account end-to-end joining of fibrils, the features of the experimentally observed kinetics are captured.

Finally, I speculate upon the consequences these observations may have on the cytotoxicity of loop-forming amyloid fibrils in the context of Chapter 3. It should be noted that it can not be said with certainty that these fibrils are amyloid in nature. While they are fibrillar and bind ThT strongly, without X-ray diffraction experiments confirming the presence of cross β-sheet structure, I will refer to these as 'amyloid-like' fibrils and not strictly 'amyloid' fibrils.

5.2 Results

5.2.1 Kinetics of Ovalbumin Fibril Formation

Fibrils were formed by reducing ovalbumin in the presence of 10mM dithiothreitol (DTT) and 10mM ammonium acetate. For initial experiments investigating the kinetics of amyloidogenesis of ovalbumin, two types of ovalbumin were used. Grade 2 ovalbumin (Sigma-A5253) is comprised of 62-68% hen egg white ovalbumin; it is essentially a crude dried egg white powder. Grade 5 ovalbumin (Sigma-A5503) is of a much higher quality with a purity $\geq$ 98%. Solutions of ovalbumin (of both grades) were pre-incubated for 2 hours at 37°C in the presence of DTT and ammonium acetate (see §2.2.2.2). After incubation the solutions were aliquotted (100 μL) into 96 well plates. The kinetics of fibril formation were measured using ThT fluorescence in a plate reader at 60°C for 24 hours. This procedure was performed for a range of initial ovalbumin concentrations (0.1 mg/ml-12 mg/ml). Representative kinetic curves for a variety of ovalbumin concentrations are shown in Fig. 5.1.

Fig. 5.1 show marked qualitative differences between 'typical' amyloid fibril kinetic curves (c.f. Fig. 3.1). Firstly, the kinetics of ovalbumin fibril formation under these conditions do not possess a discernible lag time. However, as can be seen in the traces above, there is a dip in the fluorescence at the start of the experiment. This reduction in fluorescence proceeds to a minimum value after which fluorescence increases exponentially. I define the 'flick time' as the time at which this minimum occurs (see inset to Fig. 5.5). The flick times are plotted in Fig. 5.2 as a function of concentration, for both grades of ovalbumin. As can be seen, the flick time becomes shorter with increasing protein concentration. Furthermore, flick times are effectively the same for both grades of ovalbumin.

The flick time may represent a very short lag time in the kinetics. Alter-

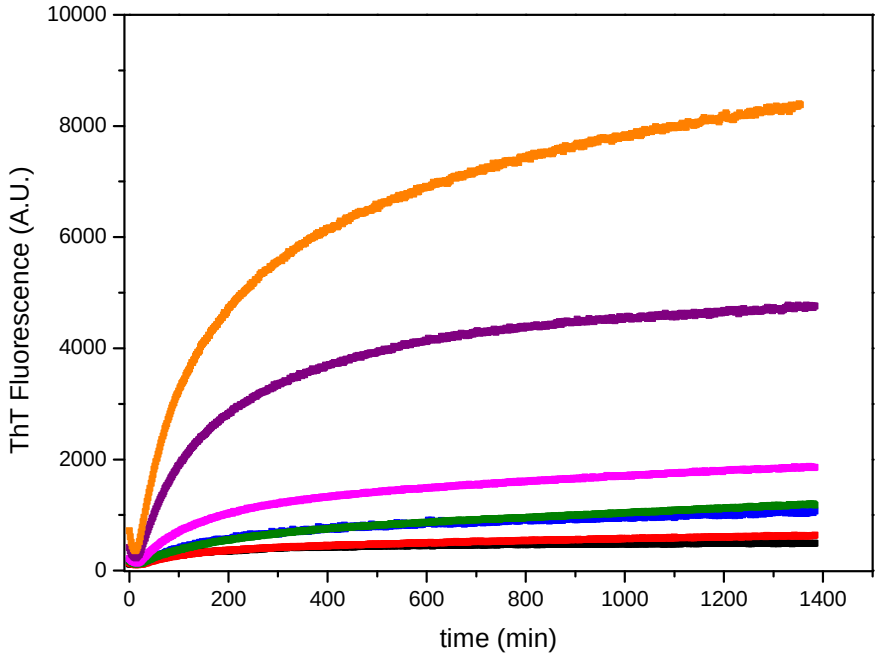

Figure 5.1: **Representative kinetic traces of ovalbumin amyloid fibril formation.** Plotted are representative kinetic traces, monitored using ThT fluorescence for Grade 5 ovalbumin: 0.1 mg/ml (black), 0.4 mg/ml (red), 0.8 mg/ml (blue), 1 mg/ml (green diamonds), 2 mg/ml (magenta), 4 mg/ml (purple), 12 mg/ml (orange).

natively, it could be an error in the way in which the fluorescence is measured. To test whether it is an artifact arising from a temperature dependence on the fluorescence signal, which may possibly be induced by heating the sample from 37 to 60°C, the flick time was determined for a constant concentration of 6 mg/ml of ovalbumin for several temperatures. These experiments were not done in the plate reader, which can only go to a maximum temperature of 60°C. I used a Cary Fluorimeter outfitted with a Peltier controlled cuvette holder. This cuvette holder allows readings up to 80°C with a much more precise level of temperature control than within a plate reader. As is shown in Fig. 5.3A, the flick time decreases with increasing temperature; by 70°C, the flick time goes to zero. It should also be noted that the fluorescence value at the minimum is roughly the same for each temperature investigated (Fig. 5.3B). I conclude that the the flick

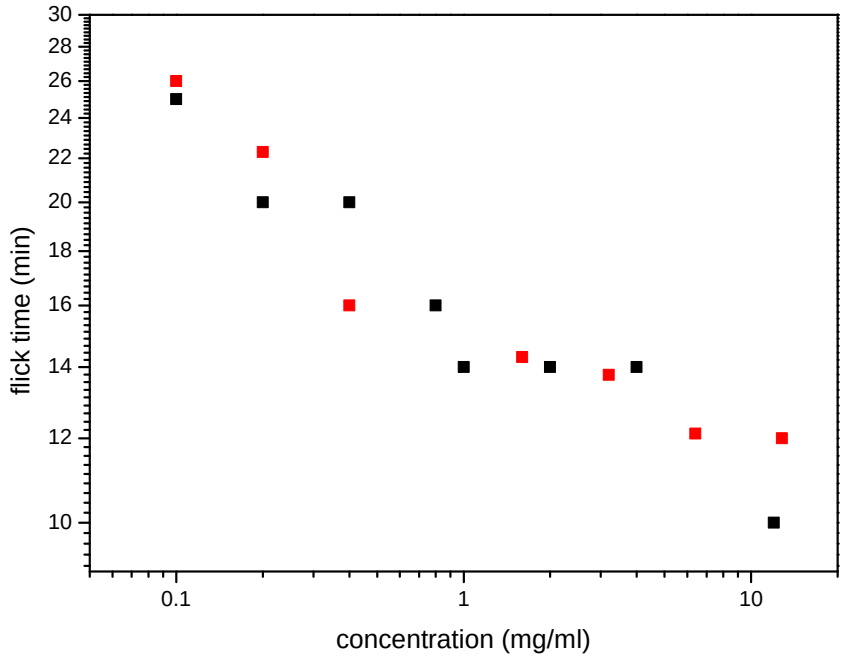

Figure 5.2: **'Flick time' as a function of concentration for two grades of ovalbumin** Plotted are the flick times for Grade 5 ovalbumin (black) and Grade 2 ovalbumin (red). In each case, the flick time decreases with concentration.

time is not in fact a proper lag time but could be the result of a decrease in the ThT quantum yield as the sample is being heated. More works needs to be done, in particular performing fluorescence lifetime measurements, in order to confirm this hypothesis. Indeed, the fact there is a non-zero fluorescence value at t=0 is indicative there are ThT-binding species already present at the beginning of the readings. Moreover, the fluorescence at t=0 is proportional to the initial concentration of ovalbumin (Fig. 5.4). Taken together, I can conclude that the kinetics show no lag time and thus resembles seeded amyloid fibril growth (§1.4.1.1; [49]). This seeding could be a result of the preparation process. The ovalbumin is reduced with DTT, breaking the internal disulphide bonds rendering the protein much more conformationally flexible and, in turn, more prone to aggregation. This susceptibility to aggregation , I hypothesize, enables the protein to be 'growth competent', i.e. act as immediate seeds for further aggregation/polymerization.

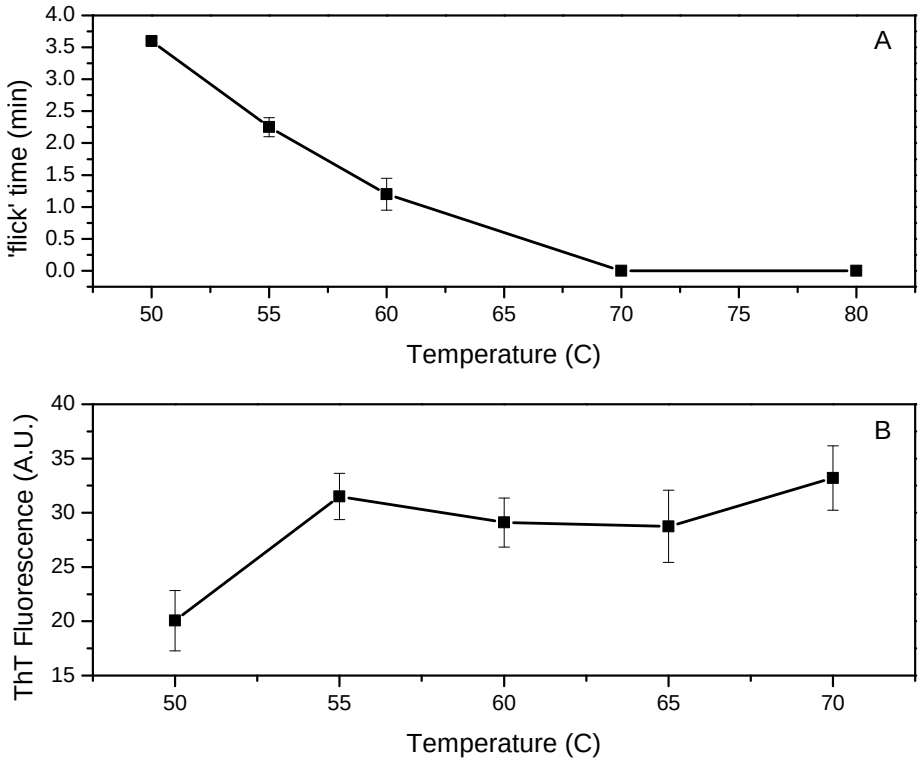

Figure 5.3: **Temperature dependence on the 'flick time'** (**A**) The 'flick time' is plotted as a function of temperature. We see that with increasing temperature, the 'flick time' decreases and may be eliminated at temperatures $> 70°$C. (**B**) The fluorescence value at the minimum is plotted as a function of temperature. I find the fluorescence value is roughly equivalent for each temperature.

After the flick time has elapsed, the kinetics of fibril assembly move into an exponential growth phase. This phase is characterized by a maximum fluorescence growth rate k_1. As exponential growth proceeds, the monomer pool begins to become depleted causing the curve to turn over as in the insulin case. However, instead of the kinetics reaching a plateau, the fluorescence slowly increases. Indeed, the data are typically well-fitted by a combination of two exponentials. We define this slow growth rate as k_2. See Fig. 5.5 for a schematic of these definitions.

A plot of k_1 as a function of ovalbumin concentration is shown in Fig. 5.6 for both grades of ovalbumin. Over the range of concentrations investigated, k_1 is linear for both grades of ovalbumin. The impure Grade 2 ovalbumin exhibits fluorescence growth rates much lower than the higher quality ovalbumin samples.

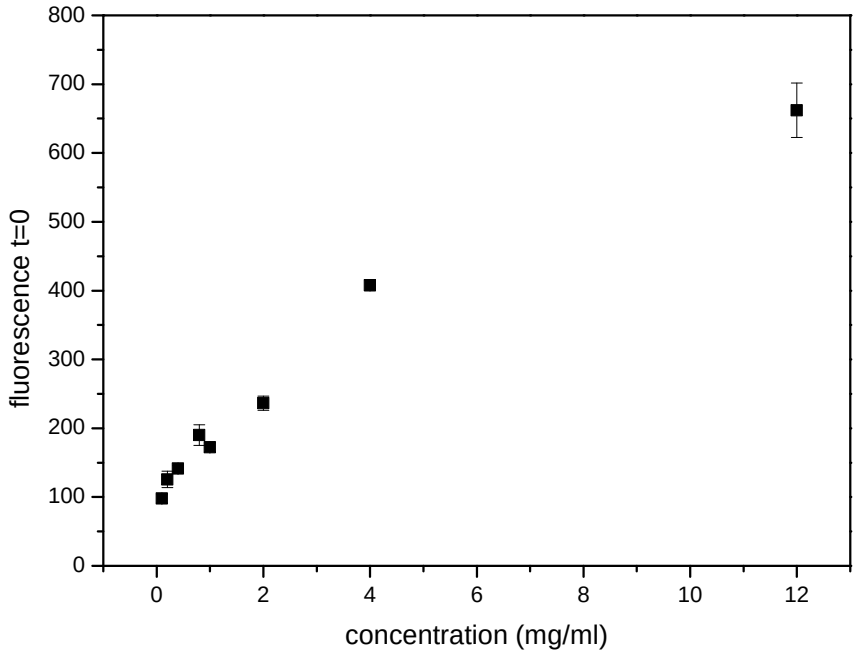

Figure 5.4: **Fluorescence at t=0 is not constant with increasing concentration.** Plot of the fluorescence value at t=0 for the concentrations investigated (Grade 5). This initial fluorescence value scales with the initial ovalbumin concentration. I attribute this to the fibril growth process being effectively seeded.

The long-time growth rate, k_2, is plotted as a function of concentration (Fig. 5.7); this trend shows a non-linear relationship between k_2 and concentration with overall lower rates for the lower grade ovalbumin samples. For the remainder of the Chapter I conducted experiments only on Grade 5 ovalbumin due to its kinetics being quicker as well as minimizing the error introduced by any additional substances (such as salts) present in the Grade 2 ovalbumin.

Again, it is useful to contrast this behaviour with the insulin amyloid fibril system studied in Chapter 3. Recall, for insulin the maximum growth rate showed a complex and non-linear dependence as a function of concentration. In that case the growth curves are normalized to remove the explicit dependence on fluorescence. This was possible due to the fact the assembly process comes to completion; the monomer pool is essentially exhausted and the kinetic curve plateaus to a steady-state value. For the case of ovalbumin, this non-linear

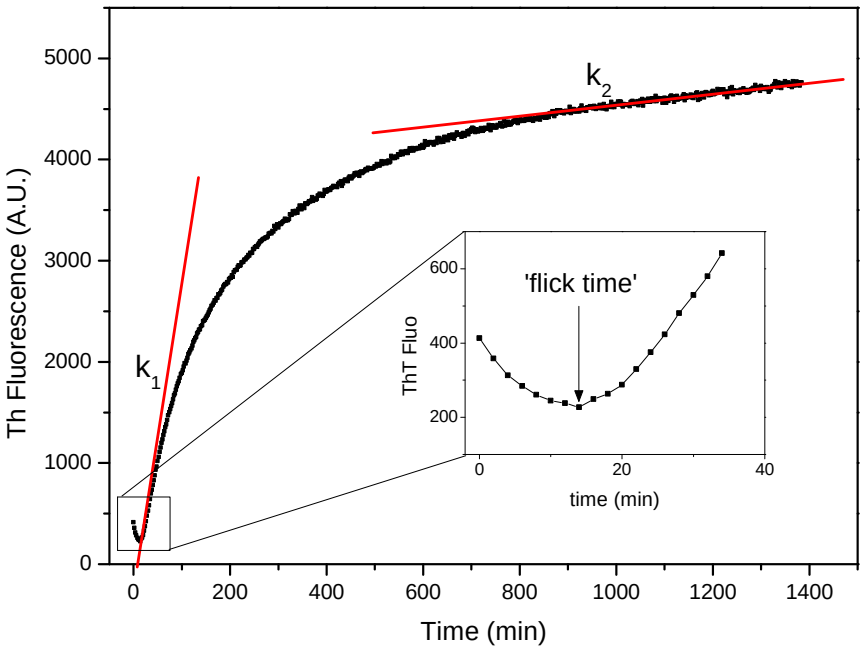

Figure 5.5: **An illustrated example of the different phases of growth observed in ovalbumin amyloid fibril formation.** This plot shows the aggregate growth, as monitored by ThT fluorescence, of a 4 mg/ml ovalbumin sample. We first observe an initial decrease in the fluorescence which results in a minimum value. After this minimum is reached the kinetics are characterized by exponential growth. The minimum in fluorescence at the early times of the kinetics is defined as the 'flick time'(inset). The exponential growth rate is called k_1. Instead of the growth kinetics reaching a constant plateau, I observe a slow growth regime which is parameterized by the growth rate k_2.

complex behaviour is not observed. Therefore, the growth rates measured for ovalbumin need not be converted from units of fluorescence min^{-1} to simply min^{-1}. However, another issue arises: since the fluorescence signal continues to increase during the time frame of the experiment, what percentage of the monomer pool has been converted to fibril after some time t in this slow growth phase?

To determine this I took the time t to be the end of the kinetic run, which for all experiments was 24 hours. Fibril samples were formed from several different initial ovalbumin concentrations and spun down in a centrifuge. The

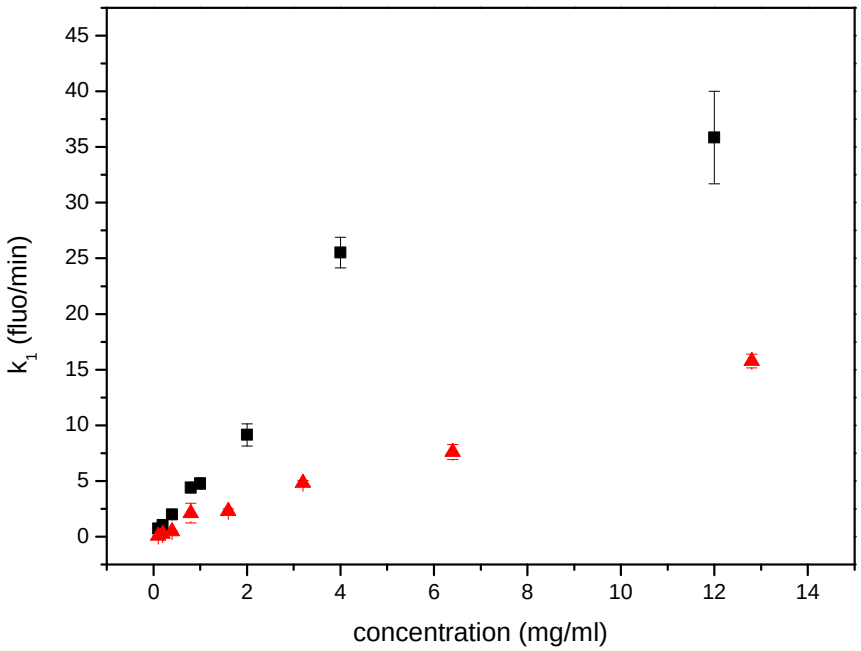

Figure 5.6: **Exponential growth rates for two grades of ovalbumin.** The growth rates during the exponential are plotted for Grade 5 (black squares) and Grade 2 (red triangles) as a function of concentration. The rates for the less pure Grade 2 ovalbumin are less than the Grade 5 possibly due to inhibiting effects of additional salt and/or proteins. Furthermore, the actual percentage of ovalbumin present in Grade 2 is roughly 70% which will render the kinetics to be slower.

supernatant was then analyzed using a Bradford Assay to determine the soluble protein concentration remaining. The results are shown in Fig. 5.8. What is immediately apparent is that after 24 hours, the fibril formation process is far from complete. Even at the 'high' concentration of 10 mg/ml, where the kinetics apparently proceed very rapidly (Fig. 5.1), only about 50% of native ovalbumin has been converted to fibrils. One consequence of this observation is that one can not simply normalize the kinetic curves to their maxima, as in the insulin case. Ultimately, this result shows that there is some mechanism which slows the depletion of the native ovalbumin population so that the polymerization reaction does not rapidly go to completion.

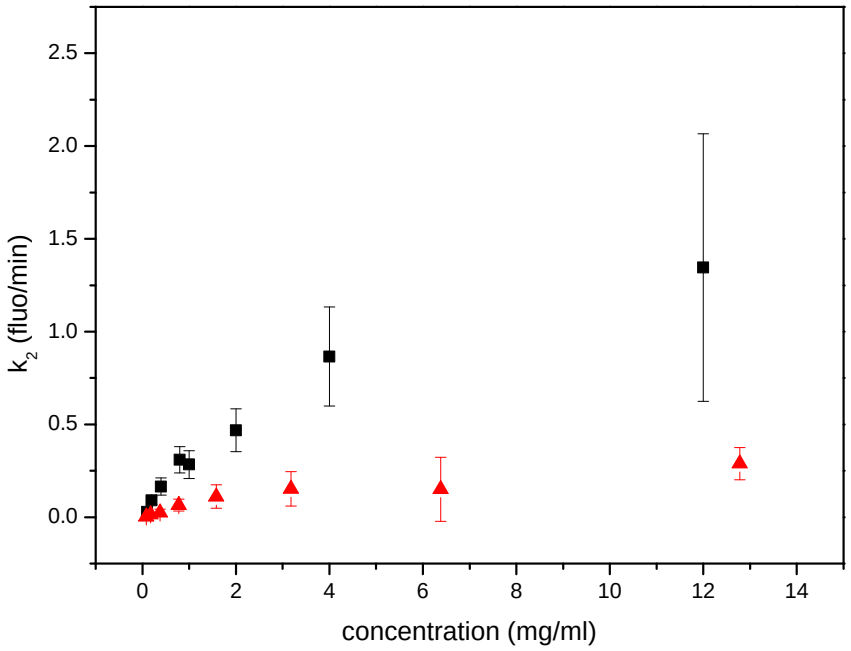

Figure 5.7: **The rates of growth in the slow growth regime for two grades of ovalbumin** The growth rate in the slow growth regime of the kinetics is shown for Grade 5 ovalbumin (black squares) and Grade 2 (red triangles) as a function of concentration. The rates for the less pure Grade 2 ovalbumin are less than the Grade 5.

5.2.2 TEM of ovalbumin fibrils

I performed negative stain TEM to study the morphological characteristics of the fibrils formed from ovalbumin under our experimental conditions (Chapter 2.2.5). As is shown in Fig. 5.9, several attributes of these fibrils are apparent. First, the fibrils are thin and apparently flexible filaments and show no detectable underlying substructure such as helical repeats. Second, images often show large entangled masses of fibrils. Third, imaged fibrils that are free from these tangled masses appear to be short (for comparison see Fig. 3.14). Finally, closed-loop fibrils are observed in many of the images (Fig. 5.9, Fig. 5.10). I define a ' closed-loop' in our images as a fibril whose ends appear to be joined to form a contiguous structure.

In order to quantify the average length and flexibility of the ovalbumin fibrils,

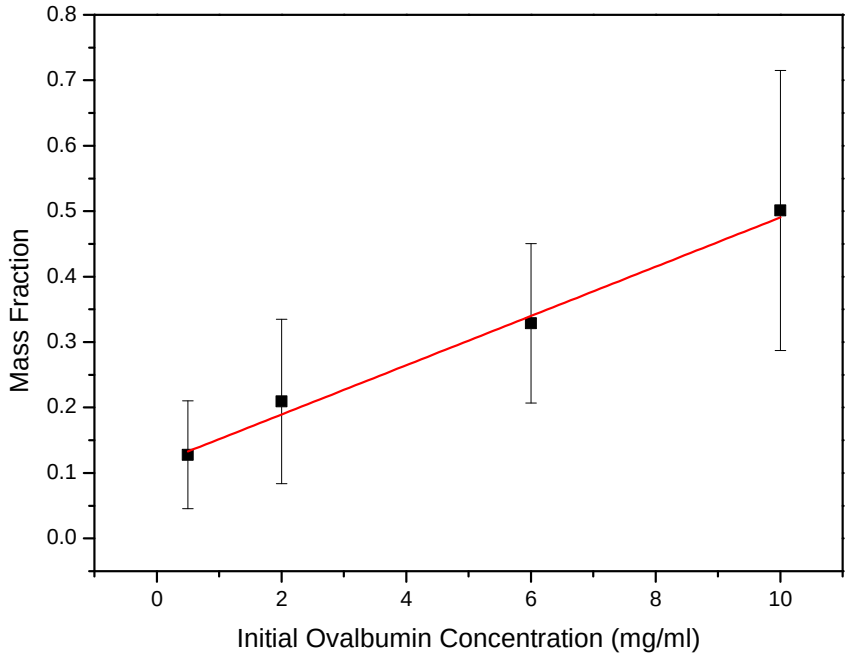

Figure 5.8: **Bradford assays show growth kinetics far from completion in the time-scale of our experiments.** We performed a Bradford Assay, using a BSA calibration standard, to determine the percentage of fibrillar material produced during the time course of our experiments. We find, even for the highest concentrations studied, after 24 hours, the fibril formation process is far from complete.

length distributions were measured and an estimate of the persistence length was made. For each image, fibrils that were clearly independent of an entangled mass and whose ends were clearly visible were chosen for measurement. Each fibril's contour length, L, was measured as well as the end-to-end distance, R. The process is shown in Fig. 5.11. I find an average value of L to be 239.8 $\pm$ 183.6 nm. The length distribution is shown in Fig. 5.12 and is well-fitted to a log-normal distribution (N=110). As ever, it can be beneficial to compare these results to what we know in the insulin amyloid fibril system. The average fibril lengths for ovalbumin are similar to fibrils formed from 'low' insulin concentrations (i.e. when $c_p < c_T$; L_{avg}=231.5 $\pm$ 230.4 nm, §3.2.5) but have a length distribution akin to when the insulin concentration is at higher concentrations ($c_p > c_T$). One conclusion I can suggest from this analysis is that fragmentation is not an active

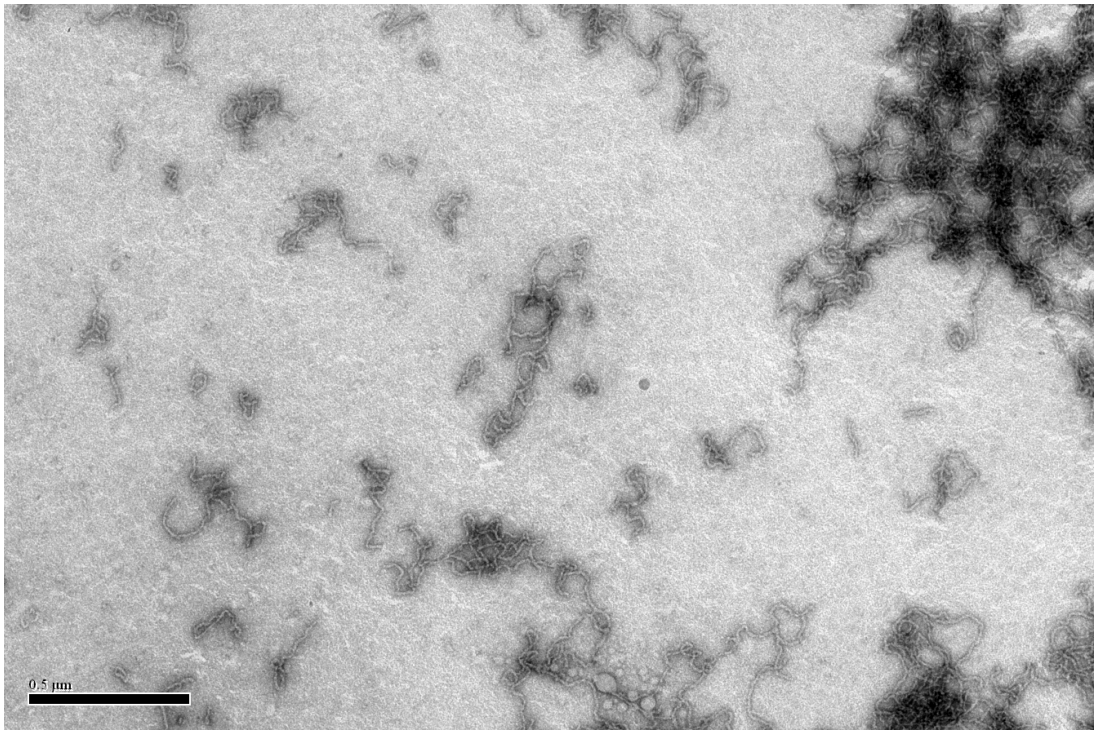

Figure 5.9: **Representative TEM image of ovalbumin fibrils.** This image is an example of a typical TEM image of fibrils formed from 6.4 mg/ml ovalbumin. They exhibit a worm-like, highly flexible morphology.

mechanism that is driving ovalbumin fibril assembly. As we saw for insulin, we would expect an exponential length distribution for a system where fragmentation is a driving force in the kinetics. The short fibril lengths can be rationalized in two possible ways. First, as has been shown, the fibril assembly process is still far from complete. One may expect that at long times fibril lengths would become longer. Second, the lack of an appreciable lag time indicates that the kinetics are in a seeded regime. One would obtain a distribution of short fibrils if the system was highly seeded so that most of the protein material is immediately growth competent. However, this scenario is less likely due to the fact that the majority of ovalbumin has not been converted to fibril after 24 hours.

The persistence length was estimated for ovalbumin fibrils by measuring the conformational parameters L and R. One must consider the effect of transferring 3-d conformation in solution to a 2-d conformation when fibrils are deposited upon a TEM grid. The conformation of a fibril observed in a TEM image can be the result of two scenarios: 1) after deposition fibril can conformationally

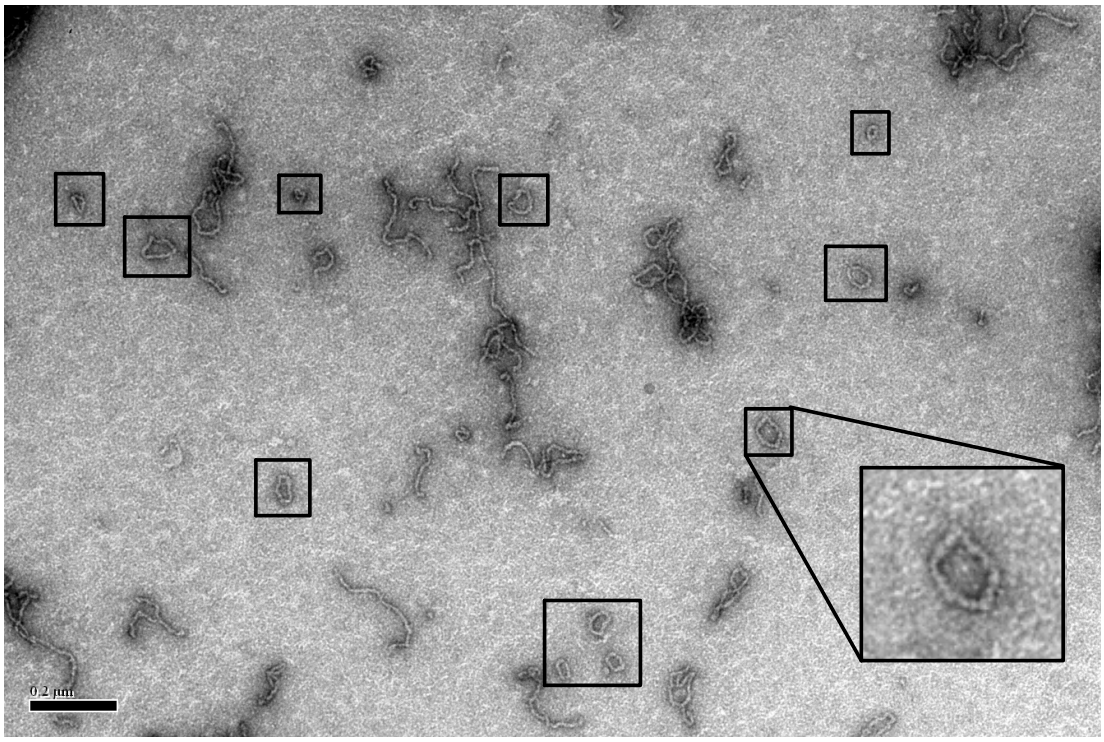

Figure 5.10: **Ovalbumin fibrils form a population of closed loop morphologies.** This is a more detailed image showing a population of closed-loop fibrils, formed from 6.4 mg/ml ovalbumin, interspersed with other worm-like fibril structures. Boxed in fibrils highlight the closed-loop morphologies. Inset shows detail of a ovalbumin fibril loop. Scale bar is 0.2μm.

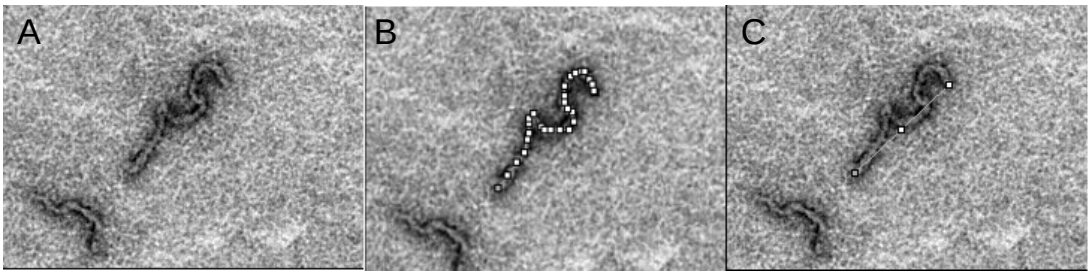

Figure 5.11: **The process of determining the contour length and end-to-end distance from TEM images** TEM images were analyzed using ImageJ software. The distance measurements were calibrated using the scale bar associated to the image. (**A**) shows an image of a fibril to be measured. (**B**)The contour length was found using the 'segmented line tool' in ImageJ; (**C**) the end-to-end distance is simply found using the 'line' tool. Fibrils were only measured where the ends were clearly discernible in the image.

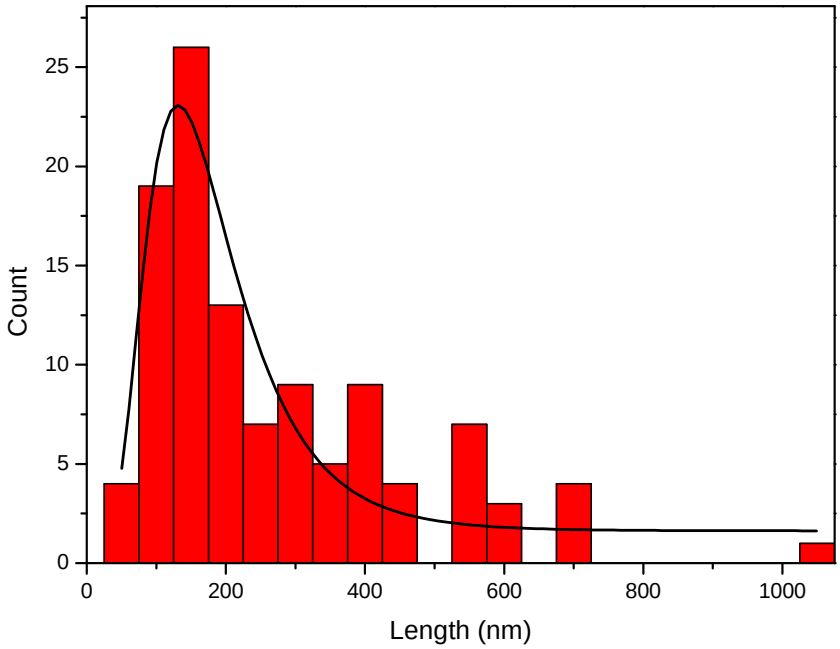

Figure 5.12: **Experimentally determined length distribution for ovalbumin fibrils.** The contour lengths of ovalbumin fibrils were found as in Fig. 5.11. The plot is a histogram of the contour lengths with N=110. The mean fibril length is 239.8 ± 183.6 nm. We find the distribution is best fit to a log-normal distribution (black curve). This result is consistent with the observation that the mean length is short due to the fibril assembly process being far from complete.

equilibrate or 2) immediately adhere to the grid surface. Furthermore, the conformations observed are assumed to arise from conformational flexibility and not inherent structural bending. It has been shown previously [204] for the case of conformationally equilibrated filaments one can obtain an equation for the mean-squared end-to-end distance, $\langle R^2 \rangle_{2D}$, of a fibril as a function of the contour length L and persistence length, P,

$$\langle R^2 \rangle_{2D} = 4PL \left(1 - \frac{2P}{L}(1 - e^{-\frac{L}{2P}}) \right) \tag{5.1}$$

It is assumed that the conformations of ovalbumin fibrils on the TEM grids are the result of conformational equilibration after deposition. R and L are

measured from our TEM images and plotted $\langle R^2 \rangle$ as a function of L. I then used the equation above to fit our data to obtain an estimate of the persistence length. The results are shown in Fig. 5.13. The persistence length is estimated to be 26 nm. For comparison, the persistence length of DNA obtained from AFM images under the same assumptions was found to be 53 nm [204]. Even more importantly, similar analysis performed on fibrils assembled from apo-CII gave an estimate of the persistence length of 36nm [56]. To compare to another fibrillar aggregate system, Smith *et al.* calculated a persistence length of 42 $\pm$ 30 μm for insulin fibrils [205]. Note the three orders magnitude difference between these two values. It may be concluded that ovalbumin fibrils are highly flexible and this flexibility helps facilitate end-joining so that closed-loop fibril conformations may be formed, as in the apo-CII case.

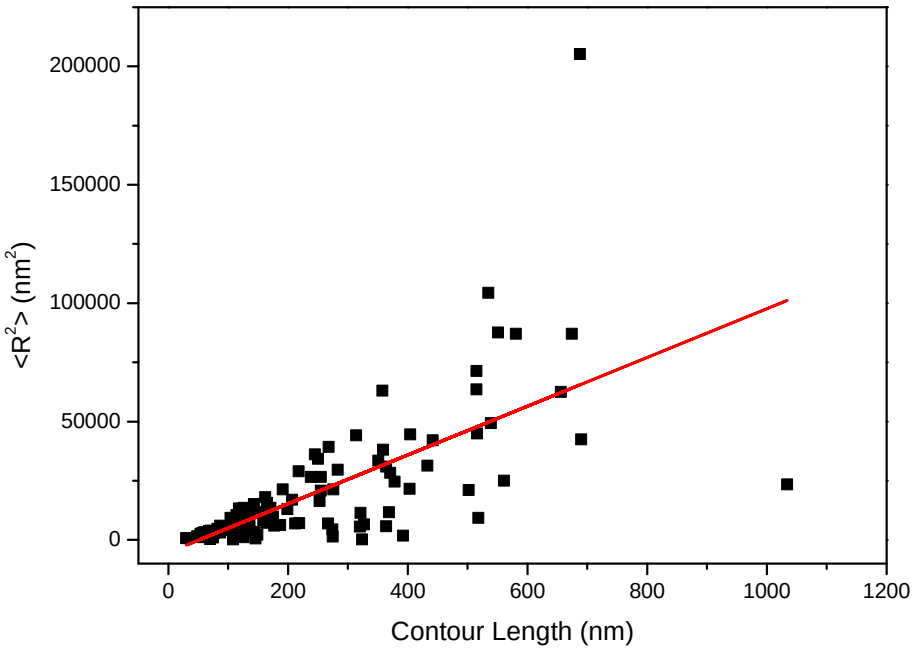

Figure 5.13: **Calculating the persistence length of ovalbumin fibrils.** Fibril contour lengths (L) and end-to-end distances (R) were found as is shown in Fig. 5.11. From this procedure R^2 is plotted against L for each fibril (N=110). The data is fit using Equation 5.1. A value for the persistence length, P, is found to be 26 nm.

5.2.3 A model of ovalbumin kinetics

I can attempt to reproduce the experimental growth kinetics in the ovalbumin system by writing down two simple models of fibril assembly, one with end-joining and one without. In the first iteration I assume a model where downhill polymerization (§1.4.1) is the dominant mechanism in the growth of the fibrils and no end-joining occurs. As we have seen from the experiments, we are in a regime where the kinetics are effectively seeded (i.e., there is no lag time) so this type of growth mechanism is plausible. We can write the time rate of change of fibrillar mass, M, with downhill polymerization as

$$\frac{dM}{dt} = k_+ E^* (m_{tot} - M(t)) \tag{5.2}$$

where k_+ is the elongation rate, E^* is the number of seeds ('ends') present at t=0, and m_{tot} is the initial mass of monomer. There is no nucleation term because we are in the seeded regime. The solution to this equation is,

$$M(t) = m_{tot} + (M_0 - m_{tot}) e^{-k_+ E^* t} \tag{5.3}$$

where M_0 is the initial seed mass. This solution is plotted in Fig. 5.14 for several different parameter values. This simple model clearly does not capture the features we observe in the experimental kinetics. First, the initial rate is too rapid. Second, there is no second slow growth phase at long times. This of course is not surprising; once the monomer pool has been exhausted the growth $M(t) \to m_{tot}$, thus reaching a plateau in the growth kinetics. If fragmentation were to be included into this simple scenario it would not ameliorate the problems; indeed, the kinetics should process even faster and still lack a slow growth phase. There is no physically realistic combination of parameters in this simple model which will yield the features I observe in our experiment: a gradual turn-over in the kinetics followed by very slow growth and a system far from coming to complete polymerization in the time scale of our experiments. Fig. 5.14 shows that the curve representing the slowest growth rate k_+ (pink diamonds) nominally reproduces the shape of the observed kinetics but fails to reproduce the polymerization time-scale I find in experiments.

As has been shown by the presence of closed loops in TEM images (Fig. 5.10, end-joining certainly occurs in this system. It would be advantageous to see how

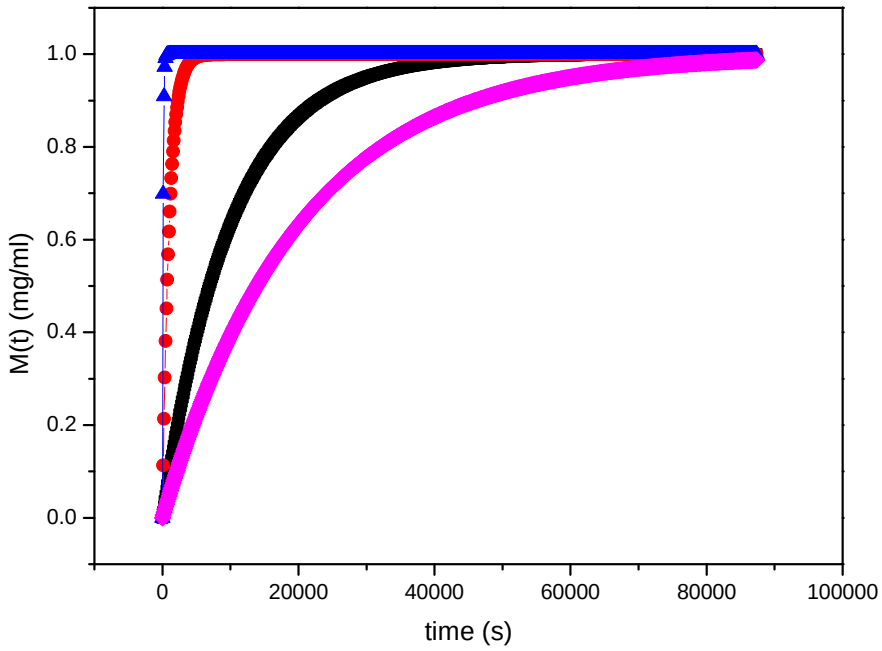

Figure 5.14: **A simple downhill polymerization model does not capture the experimental kinetics.** Plotted are solutions to Equation Fig. 5.2 for several parameters of k_+: 1x10$^{-4}s^{-1}$(blue triangles), 1x10$^{-5}s^{-1}$ (red circles), 1x10$^{-6}s^{-1}$ (black squares), 5x10$^{-7}s^{-1}$ (pink diamonds). The other parameters were held constant: M_0=1x10^{-4} mg/ml, m_{tot}=1 mg/ml, and E^*=100. For larger values of k_+, fibrils are formed extremely rapidly and the polymerization reaction quickly comes to completion. Even for small values of k_+, the amount of material converted to fibrils approaches 100% in the time window of a typical experiment. In experiments we find the percentage of fibrillar mass is far from 100% (Fig. 5.8) during the time scale of the experiments. This simple model does not match these observations. Moreover, there is an absence of the slow growth phase in this model.

adding an ending joining rate, k_j, will effect the model kinetics. I can again write down a set of equations, first for the rate of change of fibril mass M,

$$\frac{dM}{dt} = k_+ E(t)(m_{tot} - M(t)) \tag{5.4}$$

and for the rate of change of fibril ends,

$$\frac{dE}{dt} = -k_j E(t)^2 \tag{5.5}$$

where the squared term accounts for the fact there are two ends to each fibril. The solutions for M(t) and E(t) for this set of differential equations are,

$$E(t) = (k_j t + \frac{1}{E^*})^{-1} \tag{5.6}$$

and,

$$M(t) = m_{tot} - (m_{tot} - M_0)\left(\frac{1}{E^*}\right)^{\alpha} e^{-[\alpha ln(E(t)^{-1})]} \tag{5.7}$$

where $\alpha = \frac{k_+}{k_j}$.

This solution is plotted in Fig. 5.15 for several parameter values. Two features of this solution are the presence of an an initial exponential growth phase followed by a slow growth regime at longer times. Unlike in the simple downhill polymerization model where the number of ends is constant, the introduction of an end-joining rate, k_j, results in a depletion of ends over time. The decrease in the number of ends retards the rate at which monomer is converted into fibril. The net result in the long-time limit is the appearance of the slow growth phase, similar to what is observed in the experimental kinetics above. One can see from Fig. 5.15, the percentage of fibrillar mass that has been produced by the time the slow growth regime is reached is far from 100% for values of $k_j < 1\text{x}10^{-5}s^{-1}$ (with k_+=1x$10^{-5}s^{-1}$). As I have shown from the results in Fig. 5.8, the amount of fibrillar material created in our experiments is in the range of 10-60% in the time course of our experiment. I find that, for this model, such values can be nominally reproduced with k_j=1x$10^{-4}s^{-1}$ (or α=0.1).

As was shown in Chapter 3, fragmentation can act as an important mechanism for fibril assembly. One can attempt to include the effects of fragmentation into the model. However, by doing so the solution becomes non-trivial and a closed-form solution is not possible. Furthermore, the model as outlined in this thesis violates detailed balance; that is as $t \rightarrow \infty$, the number of ends $E(t) \rightarrow 0$. An incorporation of fragmentation, even if the rate is small can restore detailed balance to the model. Nevertheless, even without explicitly investigating the effects of fragmentation on an end-joining model one can posit some features of the kinetics and the resulting fibrils. For loop formation to be active, the fibrils

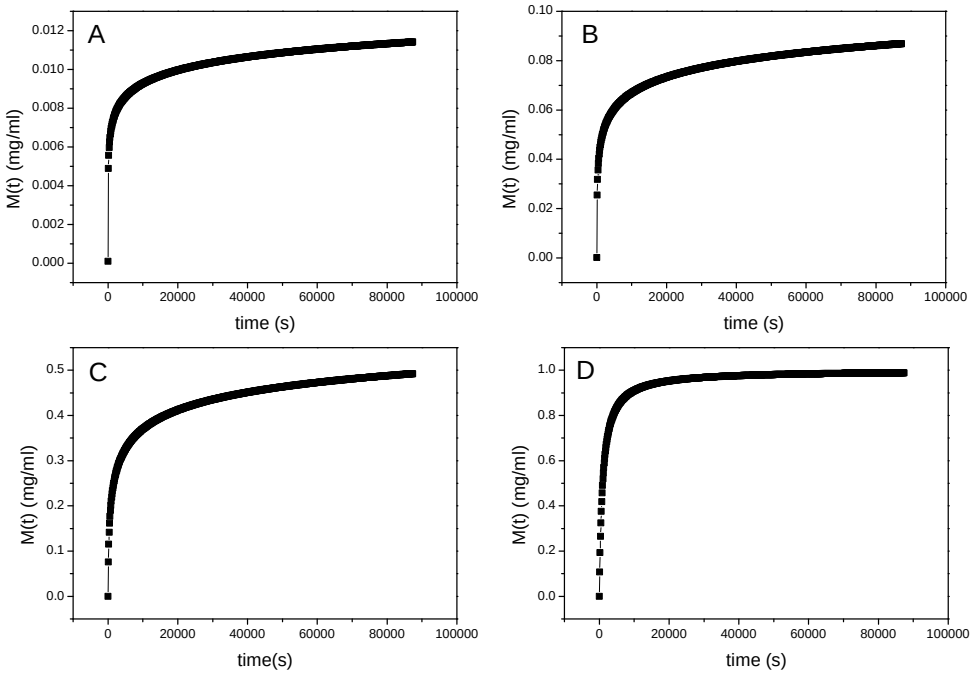

Figure 5.15: **A model with end-joining reproduces the observed kinetics.** Plotted are solutions to the set of equations Fig. 5.4 and Fig. 5.6 for several different values of the joining rate k_j: (**A**) 1x10$^{-2}s^{-1}$, (**B**) 1x10$^{-3}s^{-1}$, (**C**) 1x10$^{-4}s^{-1}$, (**D**) 1x10$^{-5}s^{-1}$. The other paremeter values were held constant: k_+=1x10$^{-5}s^{-1}$, M_0=1x10^{-5} mg/ml, m_{tot}=1 mg/ml, and $E^* = 100$. The resulting solutions show an initial exponential growth phase followed by a slow growth phase at later times. The slow growth phase exists due to the elimination of available growing ends due to joining; because of this the polymerization reaction can not come to a rapid completion. For larger values of k_j the amount of material converted to fibril is very small (~ 1 %) before the slow growth regime is reached. In our experiment we see fibril mass percentages similar to the parameter values in **C**. Note in the limit that $k_j \to 0$ the solution becomes a simple downhill polymerization.

must be flexible with short persistence lengths. Since fragmentation is postulated to be driven by thermal fluctuations [205], the intrinsic flexibility of the fibrils may result in fragmentation rates being lower than in systems with more rigid and inflexible fibrils. Second, assuming fragmentation is present, qualitatively distinct fibril length distributions would arise in systems with and without an end-joining rate. In the case where there is no end-joining the length distribution would

become exponential in the long time limit (as has been discussed in Chapter 3). Alternatively, end-joining can act as a 'self-repair' process when fragmentation is active; end-joining is not only the 'self-association' event where closed-loops are formed but fibril growth is also achieved via the concatenation of two separate fibrils into one. In other words, end-joining is the inverse process of fragmentation. In contrast, the only growth mechanism for insulin fibrils in Chapter 3 was via monomer addition only.

I hypothesized that since the length distribution is dominated by short fibrils, the resulting fibril solution should not be gel-like or highly viscous. I tested this by performing bead tracking microrheology on ovalbumin samples for a range of concentrations. As can be seen in Fig. 5.16, I found no apparent gel is formed from ovalbumin fibrils over the range of concentrations studied. The lack of formation of a highly viscous fluid is consonant with the observation that only short fibrils have been formed during the time-frame of the experiment. It is interesting to note, however, that the length distribution found experimentally is not exponential. I contrast this result to what was found for insulin (Chapter 3.2.3). Therefore, the mechanism which prevents the collapse of the fibril length distribution in ovalbumin may be distinct from that of insulin (see Chapter 3).

5.3 Discussion & Conclusion

In this chapter I have shown that the kinetics of ovalbumin fibril formation is typified by 'non-standard' assembly kinetics. Specifically, the kinetics appear seeded and lack a lag phase. It would be advantageous in the future to test this hypothesis by attempting to remove Moreover, there exists a slow growth phase which occurs after the initial exponential growth phase. The presence of this slow growth is in contrast to what typically happens in amyloid fibril assembly where a plateau is reached due to the exhaustion of competent growth species. I observed worm-like and highly flexible fibrils with a persistence length of 26nm. Interestingly, I noted the presence of closed-loop fibrils. I attribute the observed kinetic features, particularly the presence of the slow growth phase, on the ability for these fibrils to join end-to-end (in contrast to other models that explicitly assume growth by monomer addition). This end-joining mediates the formation of closed-loops which gives rise to end depletion. By writing down a simple model

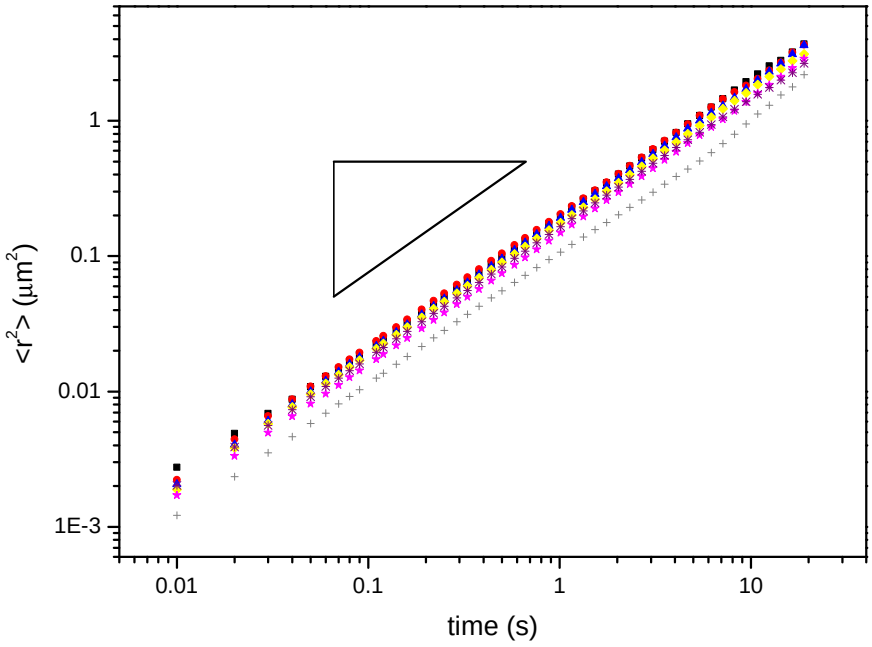

Figure 5.16: **Gels do not form from ovalbumin amyloid fibril samples as shown by bead tracking experiments** 1 μm beads were optically tracked while diffusing in amyloid fibril solutions made from ovalbumin. A range of concentrations were explored: buffer control/no fibrils (black squares), 0.375 mg/ml (red circles), 0.75 mg/ml (blue triangles), 1.5 mg/ml (yellow diamonds), 3 mg/ml (magenta stars), 6 mg/ml (purple asterisks), 12 mg/ml (grey crosses). With increasing concentration the solutions become increasingly viscous, but no viscoelastic plateau is observed. The hypotenuse of the black triangle has a slope of 1, and is provided for comparison.

with an end-joining rate, the qualitative features of the kinetics may be captured.

A natural question is how universal is the ability for amyloid fibril systems to form closed-loop morphologies. The kinetics of such a process has not been widely studied to date. However, closed-loop amyloid fibrils have been observed in a number of protein systems besides ovalbumin and apo-CII, for instance αB-crystallin [206], β_2-microglobulin [52], and α-synuclein [207]. If fibril flexibility is a good indicator for the potential of forming closed loops, many other protein systems exhibit flexible, worm-like fibrils under the correct conditions (e.g. BSA [208], Aβ(1-40) [108]). However, fibril flexibility may be a necessary but not

sufficient condition to form closed loop fibrils. Other considerations may play a role such as fibril structure or electrostatic self-interactions.

Of note is that many of these systems may exhibit many different morphologies depending on the conditions in which they were formed. For instance, Gosal and co-workers have shown in the β_2-microglobulin system, one can obtain a myriad of structures, from long straight fibrils, to flexible worm-like ones depending on the pH and ionic strength of the solution [52]. The existence of closed loop amyloid fibrils is simply an interesting addition to the amyloid fibril morphological 'zoo'. The fact that, for a given protein system, morphologies may be tuned simply by changing the conditions of formation holds interesting possibilities for creating novel amyloid-fibril based biomaterials. To this end more studies need to be done in order to better understand the conditions which promote the formation of closed-loop amyloid fibrils and how this particular morphology is connected to the other morphological pathways of fibril assembly.

This potential control of morphologies could be of great interest in regards to the potential cytotoxicity of closed-loop amyloid fibrils. As was discussed in Chapter 3, oligomeric and flexible protofibrillar amyloid species have been implicated as being the cytotoxic agents in many diseases (e.g. [186]). Lashuel *et al.* showed that mutations which cause Alzheimer's Disease (Aβ(1-40)) and Parkinson's Disease (α-synuclein) promote the formation of annular, pore-like morphologies of amyloid protofibrils. Other pore-like annular structures have been observed for other amyloid fibril forming proteins [42]. In the absence of fragmentation, a closed-loop amyloid fibril is non-reversible. The fact that loops are a kinetic end point could have implications for the prevalence and toxicity of these species.

Conversely, if it were the case that larger, non-pore forming closed-loop amyloid fibrils were not cytotoxic, forming loops may prove to be beneficial. In a system where fragmentation creates short, oligomeric species, the formation of loops may act as a kinetic sink, stabilizing the length distributions and preventing these shorter and potentially toxic species from accumulating. This mechanism that prevents run-away increase in potentially harmful oligomeric species is in contrast to the physical arrest of fragmentation which was discussed in detail in Chapter 3.

Chapter 6

Investigating ocr aggregation by studying the gelation of genetically and chemically modified variants.

6.1 Introduction

6.1.0.1 Ocr Gelation

In a typical ammonium sulphate precipitation step during protein purification, less than 2M ammonium sulphate is used to 'salt out' protein from solution. However, it was observed that at high ammonium sulphate concentrations ($\sim$ 3M), instead of precipitating, ocr formed a weak gel. Richard Higham published a PhD thesis elucidating some of the physical and phenomenological aspects of ocr gelation [4]. Curiously, ocr gelation seems to be salt specific. Gelation of the wild-type, native protein only occurs in the presence of a critical concentration of ammonium sulphate. In Higham's thesis, despite using salts with similar properties to ammonium sulphate, gelation was never observed. In addition, AFM results showed that ocr forms fibrils several nanometers thick and with lengths over 1 μm (Fig. 6.1). Such high charge-screening conditions usually result in particulate gels or protein precipitates (see Chapter 1.2.1). The fact that ocr, under these experimental conditions, forms a 'fine-stranded' translucent

gel is, to the best of my knowledge, unprecedented.

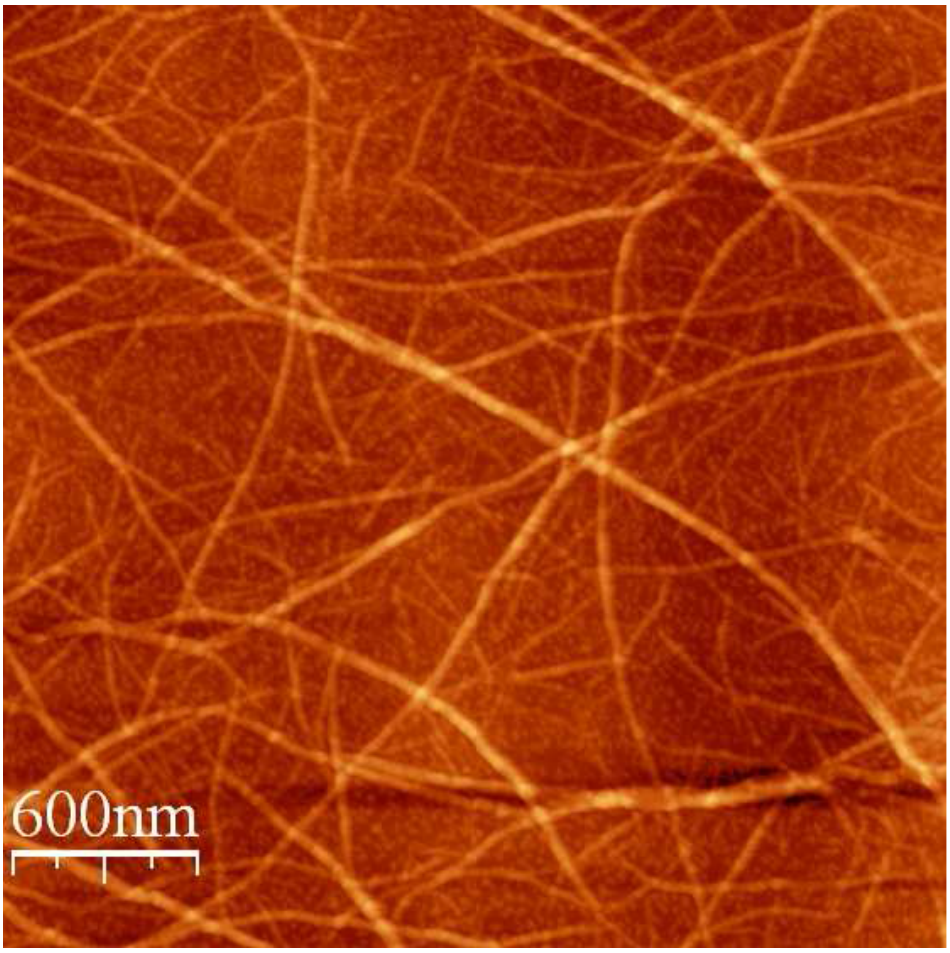

Figure 6.1: **AFM images of ocr fibrils** AFM image of 36 μM ocr in 20 mM Tris-HCl pH 8 with 3.2M ammonium sulphate. Typical dimensions of fibrils from AFM measurements show heights $\sim$ 10 nm with lengths up to 1 μm. Taken from [4].

There are many open questions about the gelation of ocr. Most pressing is the mechanism of ocr fibril formation. For instance, unlike the aggregation of protein into amyloid fibrils, there appears to be no conversion of the native secondary structure into a significant proportion of β-sheet. Furthermore, ocr gelation is a reversible process; cycling through aggregation and dis-aggregation does not appear to have ill effects on the native conformation. In this chapter, we attempt to gain a better understanding of ocr aggregation by studying the gelation of ocr mutants. These mutant variants each contain modifications to

regions containing high surface charge density. I use bead tracking microrheology to study the structural characteristics of the gels formed from these mutants in order to identify important regions on the surface which contribute to ocr's aggregation behavior.

6.2 Methods

6.2.1 Ocr Variants

Mutant variants of ocr were produced using genetic modification so that negative charge regions on the surface of the protein were neutralized. More specifically, acidic aspartates and glutamates were mutated to asparagines and glutamines, respectively. The details of the genetic modification protocol can be found in [209]. In this chapter we studied three mutants: 'ocr-Mutant7' (Mut7),'ocr-Mutant11' (Mut11), and 'ocr-pOcr'(pOcr). The amino acid modifications for each are:

Mut7:

D25C, D26N, D29N, D31N, D32N, D35N, D92N, E95Q, D96N, E98Q, D99N

Mut11:

D25C, D26N, D29N, D31N, D32N, D35N, E59Q, D62N, E64Q, E66Q, D67N

pOcr:

D76N, E87Q, D92N, E95Q, D96N, E98Q, D99N, E103Q, E106Q, E107Q, E109Q, E110Q, E112Q, E113Q, E114Q, E115Q, E116Q

Note that Mut7 and Mut11 share a region of common modification. The majority of pOcr modifications take place toward the C-terminal end of the protein which is comprised of both α-helix and unstructured coil. Diagrams of these modifications, visualized for a dimeric ocr are shown in Figs. 6.2, 6.3, and 6.4.

Additionally, I also studied a truncated variant of ocr (ocr99), where the final 17 amino acids (10 of which are acidic residues) were chemically excised.

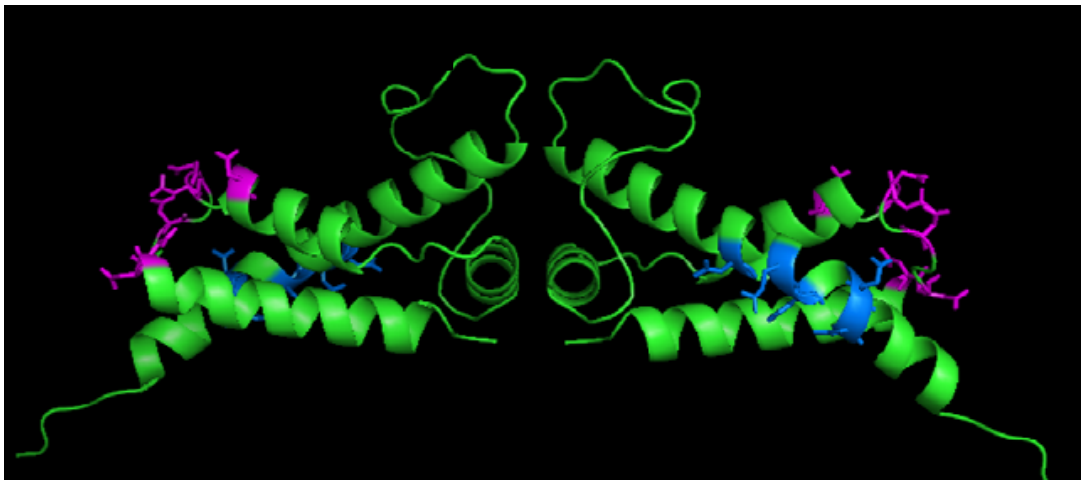

Figure 6.2: **Diagram of 'ocr-Mutant7'** Regions of genetic modifications of the ocr dimer are highlighted. Amino acids substitutions are: D25C, D26N, D29N, D31N, D32N, D35N (magenta); D92N, E95Q, D96N, E98Q, D99N (blue).

6.3 Results

We performed bead-tracking microrheology experiments on wt-ocr and the mutant variants (see §2.2.6.4 for details on the experimental methods). We first performed bead-tracking microrheology on wild-type ocr to establish a baseline for comparison with the modified ocr variants. The mean squared displacements (MSD) for a range of ammonium sulphate concentrations are shown in Fig. 6.5. I find that from 0 to 2.4M ammonium sulphate, the samples become increasingly viscous. I then find a sharp change in the scaling of the MSD at 2.8M. By 3.3M the scaling exponent has decreased to a value 0.12. This value indicates that a viscoelastic solid has formed.

Looking first at the raw data from microrheology we can infer different behaviour for the three mutant proteins. Mut7 begins to display increased viscosities at 0.8M ammonium sulphate, similar to what was found with wt-ocr (Fig. 6.6). However unlike wt-ocr, at 1.6M we find an abrupt change in the MSD scaling behavior. With further increasing ammonium sulphate concentrations, the MSD scaling relation continues to decrease. Interestingly, the MSDs for Mut11 are remarkably similar to that of Mut7 (Fig. 6.7). Finally, pOcr exhibits markedly different behavior from wt-ocr and the other mutant variants. As can be seen in Fig. 6.8, the slope of the MSD goes to zero with only a 0.8M

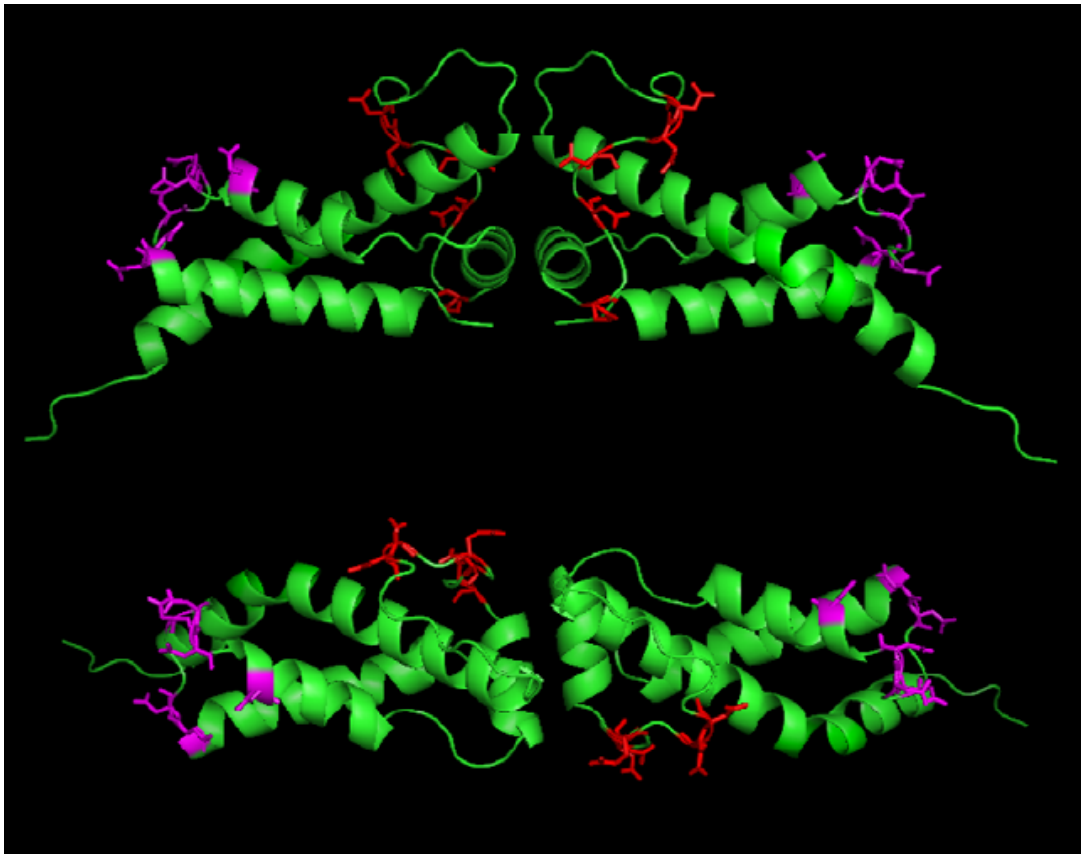

Figure 6.3: **Diagram of 'ocr-Mutant11'** Regions of genetic modifications of the ocr dimer are highlighted. Amino acids substitutions are: D25C, D26N, D29N, D31N, D32N, D35N (magenta); E59Q, D62N, E64Q, E66Q, D67N (red). The bottom image of the ocr dimer is rotated by 90° relative to the top view.

ammonium sulphate concentration, indicating a gel state is reached. Further increasing the ammonium sulphate concentration does not dramatically effect the MSD scaling. This marked change in the gelation behaviour may begin to highlight an important region on ocr, namely the final $\alpha-helix$ and unstructured tail on each monomer of ocr.

To better visualize the changes taking place for each variant of ocr, the viscosities are plotted as a function of ammonium sulphate concentration. Viscosities were calculated from the MSDs and an application of the Stokes-Einstein equation (§2.2.5). As is discussed in §2.2.5, we assume that at in the long time limit, the scaling exponent for all MSDs goes to one but with a much reduced diffusion coefficient relative to pure water. For wt-ocr (Fig. 6.9a), we find

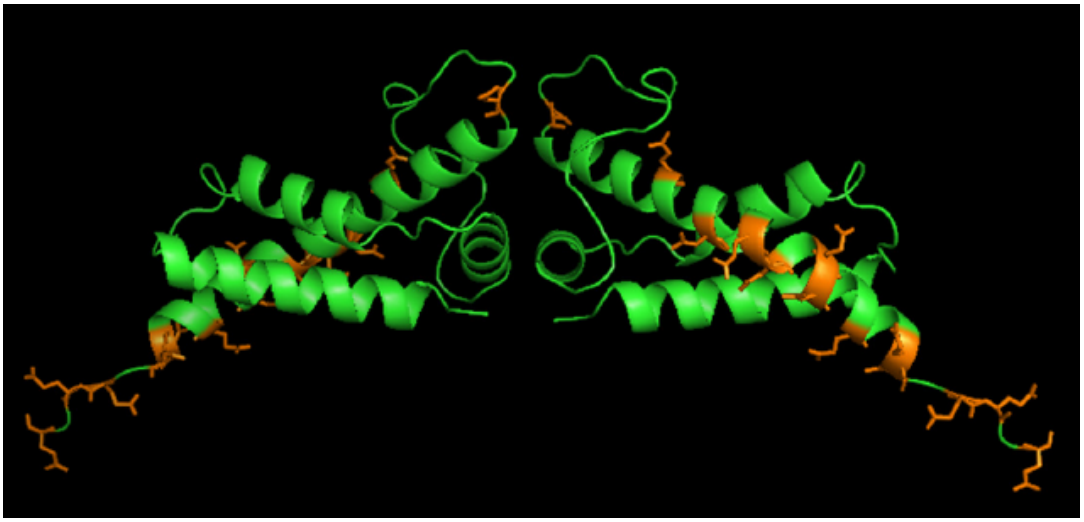

Figure 6.4: **Diagram of ocr-POcr Mutant** Regions of genetic modifications of the ocr dimer are highlighted. Amino acids substitutions are: D76N, E87Q, D92N, E95Q, D96N, E98Q, D99N, E103Q, E106Q, E107Q, E109Q, E110Q, E112Q, E113Q, E114Q, E115Q, E116Q.

a gradual increase in the viscosities up to 2.4M. At 2.8M we find a sharp increase in the viscosity. Further increasing the ammonium sulphate concentration to 3.3M results in a further increase of the viscosity to a value of 541 cP. Mut7 and Mut11 again show a steady increase in viscosity from 0-1.2M (Fig. 6.9b,c). At 1.6M we find an appreciable increase in the viscosity. Again, increasing the ammonium sulphate concentration results in an additional increase in the viscosity up to final values of 524 and 595 cP at 3.3M, respectively. In the case of pOcr, the sharp increase in viscosity occurs at 0.8M (Fig. 6.9d). Additional increases of ammonium sulphate concentration do not lead to an appreciably different viscosity. Furthermore, the average viscosity for pOcr (0.8-2.8M) is 311.7 $\pm$ 45.3 cP which is less than wt-ocr and the other mutant variants. An interesting feature that is found in Fig. 6.9 is that there is an increase in the viscosity at lower ammonium sulphate concentrations for wt-ocr (1.2M) and the two mutants, Mut7 and Mut11 (0.8M). This feature may correspond to the formation of higher order aggregates or the assembly of filaments without complete entanglement.

A final ocr variant was studied, namely, the truncated version of ocr where the final 17 amino acids have been chemically excised (ocr99). The experiments involving this truncated ocr variant were performed early in the process of

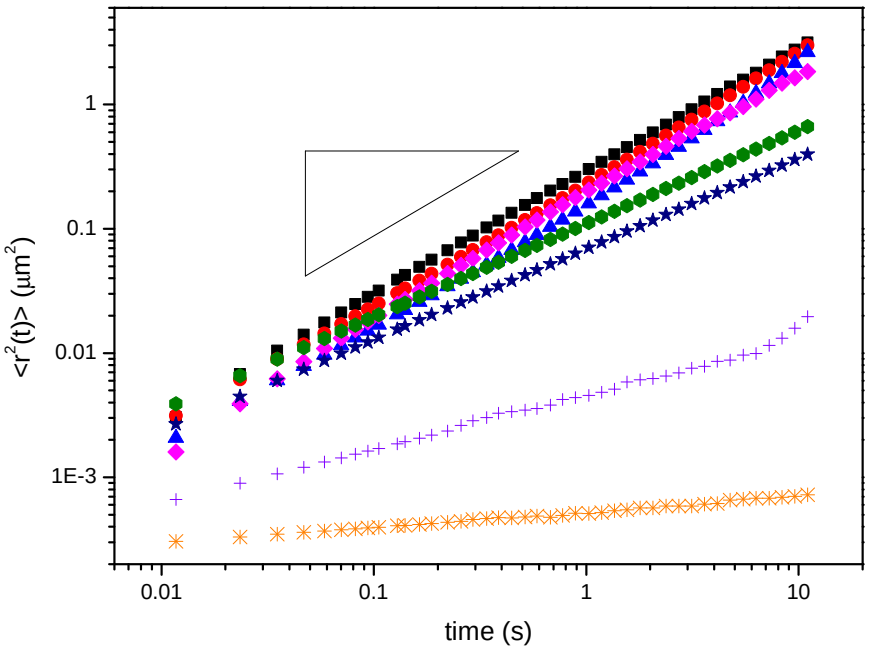

Figure 6.5: **Mean squared Displacements for wild-type ocr.** Plot of the MSDs for a range of ammonium sulphate concentrations (M): 0 (black squares), 0.8 (red circles), 1.2 (blue triangles), 1.6 (pink diamonds), 2 (green hexagons), 2.4 (navy stars), 2.8 (purple +), 3.3 (orange *). Triangle is a guide for the eye; the hypotenuse shows a slope of 1 which is equivalent to a liquid with the viscosity of water.

developing the methods employed in this Chapter. A lower frame capture rate (10 fps) was used and only two movies were made with only of order $\sim$ 10 beads per movie. Furthermore only a reduced range of ammonium sulphate concentrations were investigated however there was insufficient protein to complete these experiments. However, the data will still be presented for while it may not allow for reliably quantitative comparisons with the other variants, qualitative ones may still be made. Between 0.8M and 2M there is an increase in the viscosity. At 2.4M we find a sharp change in the scaling of the MSD corresponding to the formation of a visco-elastic gel.

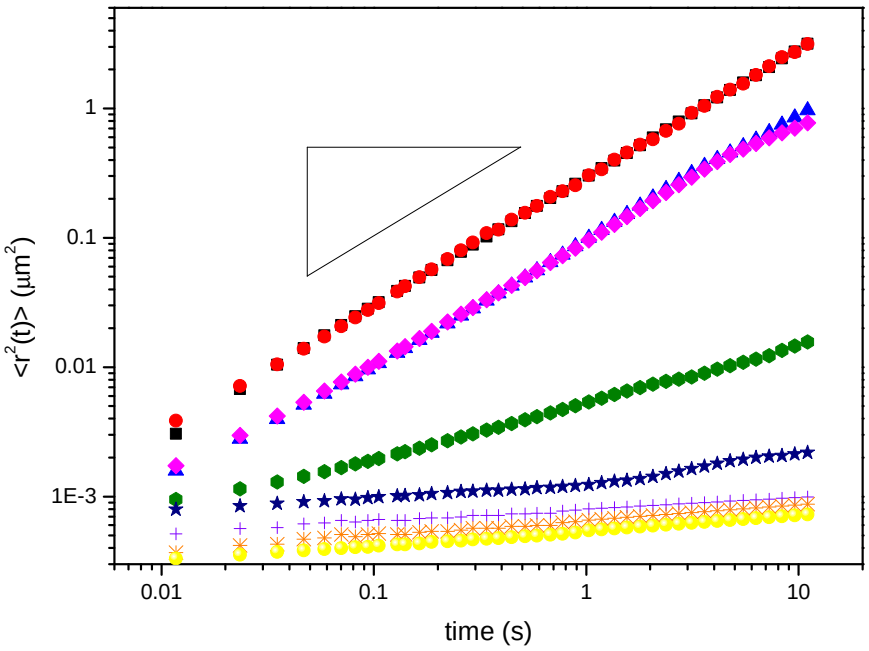

Figure 6.6: **Mean squared Displacements for Mut7 ocr.** Plot of the MSDs for a range of ammonium sulphate concentrations (M): 0 (black squares), 0.4 (red circles), 0.8 (blue triangles), 1.2 (pink diamonds), 1.6 (green hexagons), 2 (navy stars), 2.4 (purple +), 2.8 (orange *), 3.3 (yellow spheres). Triangle is a guide for the eye; the hypotenuse shows a slope of 1.

6.4 Discussion & Conclusion

This chapter has attempted to elucidate the mechanism of ammonium sulphate induced aggregation of ocr by studying the visco-elastic properties of the variants (both genetically and chemically modified) using microrheological techniques. From this work several conclusions may be drawn. First, the modified variants of ocr exhibit markedly different aggregation behavior from the wild-type as a function of ammonium sulphate concentration. The concentration of ammonium sulphate at which the variants will form a gel is less than the wild-type for all variants studied. This result may be rationalized by the fact that the decreased charge allows for smaller ammonium sulphate concentrations to effectively shield the electrostatic interactions and induce aggregation. Indeed, the gel-point ammonium sulphate concentration does scale well with surface charge (Fig. 6.11).

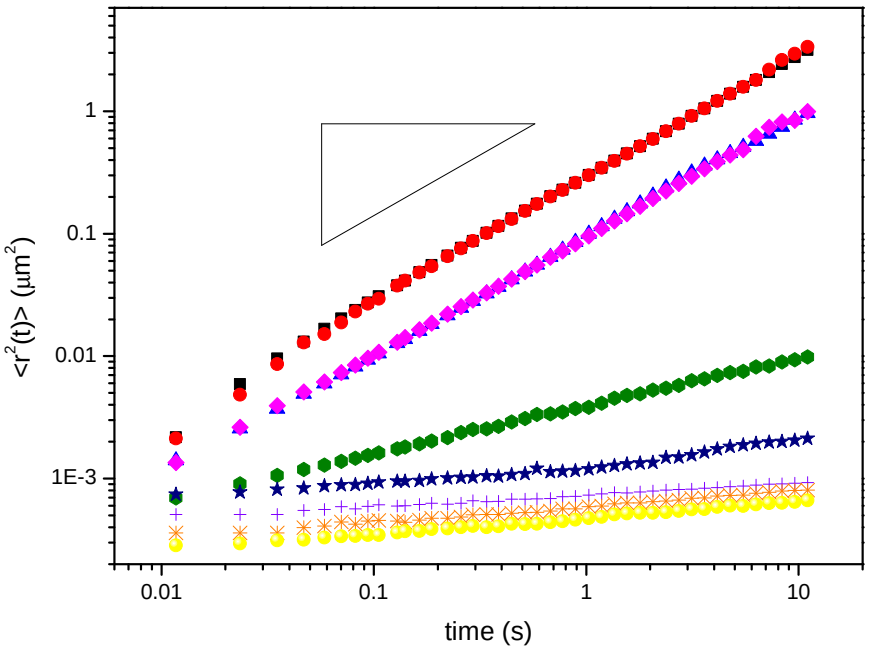

Figure 6.7: **Mean squared Displacements for Mut11 ocr.** Plot of the MSDs for a range of ammonium sulphate concentrations (M): 0 (black squares), 0.4 (red circles), 0.8 (blue triangles), 1.2 (pink diamonds), 1.6 (green hexagons), 2 (navy stars), 2.4 (purple +), 2.8 (orange *), 3.3 (yellow spheres). Triangle is a guide for the eye; the hypotenuse shows a slope of 1.

Second, it has been shown that the aggregation behavior is very similar for Mut7 and Mut11. Both variants contain an overlap in modification, namely the substitutions D25C, D26N, D29N, D31N, D32N, D35N. Mut7 and Mut11 have further modifications in different regions of the protein (see Figs. 6.2 and 6.3). One possibility is that the aggregation behavior and structural characteristics of the resulting gel are so similar because the shared region of charge modification has a greater impact on the aggregation of ocr than the other regions of modification of these mutants. Alternatively, the phenomena observed is simply a result of electrostatics, regardless of which amino acid side chains are changed.

Third, pOcr had the lowest ammonium sulphate concentration gel point of all the variants. While the ammonium sulphate concentration gel-point may be interpreted in terms of a loss of surface charge due to amino acid charge

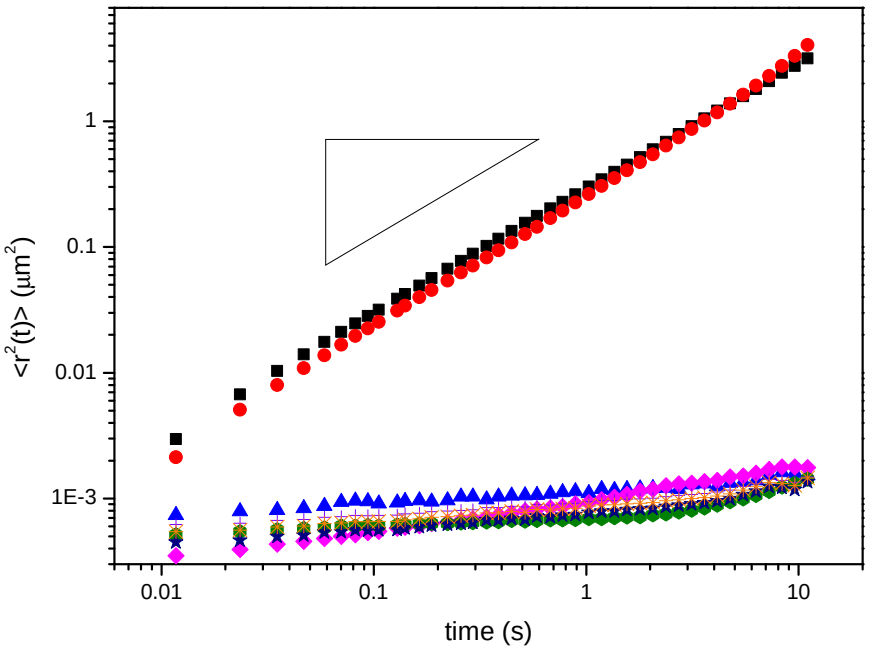

Figure 6.8: **Mean squared Displacements for pOcr ocr.** Plot of the MSDs for a range of ammonium sulphate concentrations (M): 0 (black squares), 0.4 (red circles), 0.8 (blue triangles), 1.2 (pink diamonds), 1.6 (green hexagons), 2 (navy stars), 2.4 (purple +), 2.8 (orange *). Triangle is a guide for the eye; the hypotenuse shows a slope of 1.

substitutions, the fact the gel is physically weaker than the wild-type and other genetically mutated variants is of interest. This could be an indication that the regions of modified charge on pOcr could strongly influence the formation of aggregates and their structural integrity.

Fourth, ocr99 does not exhibit similar aggregation behavior as pOcr. The excised portion missing from ocr99 contains 10 of the 17 charge modifications of pOcr (predominately located in the unstructured C-terminus tail). We find the gel points of pOcr and ocr99 occur at 0.8M and 2.4M, respectively. One can conjecture that the excised portion of ocr99 is not a critical region of charge that influences the aggregation behavior. The fact that the ammonium sulphate concentration gel-point of ocr99 is lowered does indicate it does play some role; the loss of the charge in that region results in less non-specific electrostatic repulsion

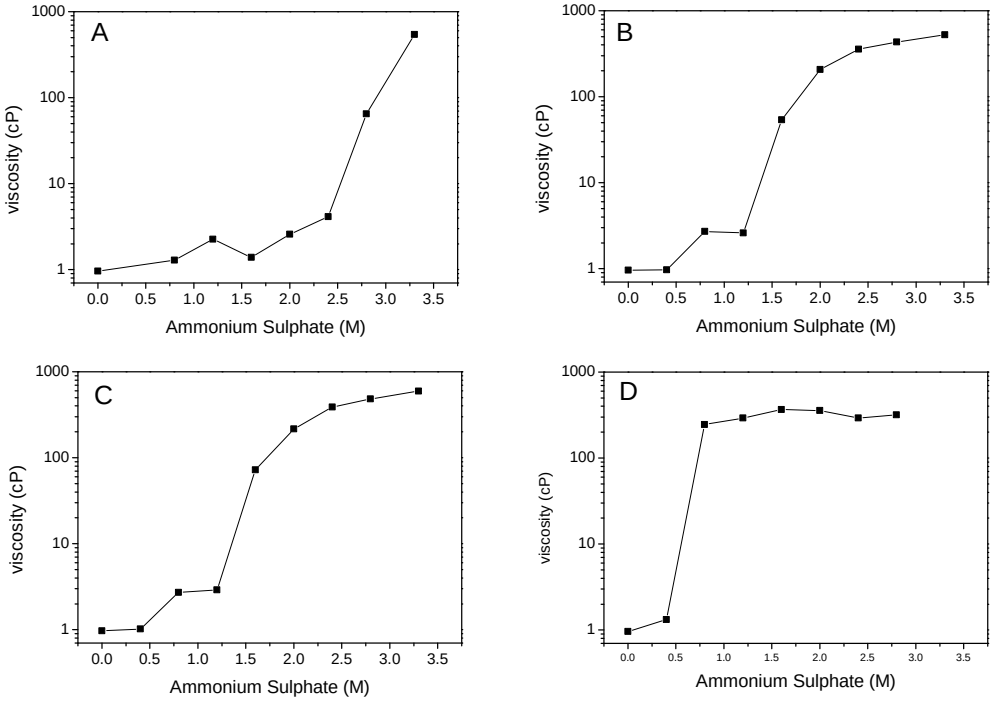

Figure 6.9: **Viscosities of ocr and its mutant variants as a function of ammonium sulphate concentration.** Plotted are the viscosities determined from the mean squared displacements and applying the Stokes-Einstein equation: (**A**) wt-ocr, (**B**) mut7, and (**C**) mut11, and (**D**) pOcr. pOcr behaves distinctly from wt-ocr and the Mut7 and Mut11: there is a transition to a gel-like state at 0.8M ammonium sulphate. Moreover the gel formed is apparently weaker than the other samples tested. Mut7 and Mut11 exhibit very similar viscosities as a function of ammonium sulphate concentration.

between proteins. However, the regions of charge where the modifications of ocr99 and pOcr that do *not* intersect (the amino acids D76, E87, D92, E95, D96, E98, D99) indicates they have a stronger role in the aggregation behavior of ocr. The location of these amino acids are primarily located in the last α-helix prior to the C-terminal (on each monomer unit). The aggregation behavior of pOcr, which containing this modified region, may indicate that this α-helical region and the charge present there may be a specific region on the protein surface which contributes strongly to the gelation behavior of ocr. Moreover, while the surface charges of ocr99 and the mutants Mut7 and Mut11 are similar, the ammonium sulphate concentration where the gel point is observed is different

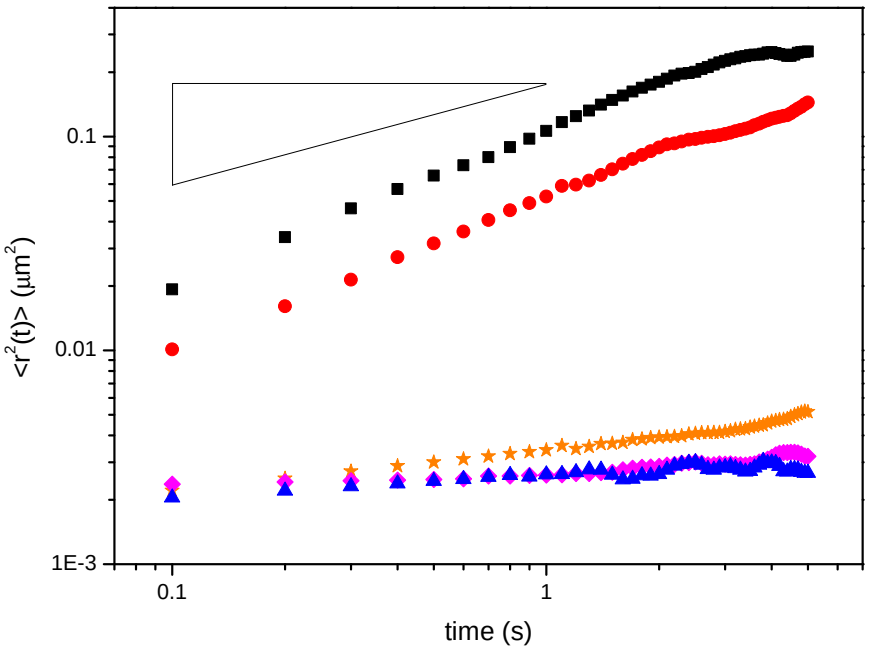

Figure 6.10: **Mean squared displacements for ocr99.** Plot of the MSDs for a range of ammonium sulphate concentrations (M): 0.8 (black squares), 2 (red circles), 2.4 (blue triangles), 2.8 (pink diamonds), 3.3 (orange stars). Triangle is a guide for the eye; the hypotenuse shows a slope of 1.

for ocr99 and Mut7/11(2.4M and 1.6M, respectively). This is further indication that this specific truncated portion of ocr is of less importance than other regions of surface charge.

The question remains to be answered: what is the mechanism which causes ocr to form fibrillar aggregates at high ammonium sulphate concentrations? (Fig. 6.11) follows an archetypal salting-out trend, yet ocr and the variants studied do not precipitate. It was noted in [4] that at pH values $\sim$ 4, ocr *will* precipitate. Theoretical titration curves predict a surface charge of $\sim$ -30e at pH 4. It is interesting to note that pOcr, which forms a weaker gel at a much lower ammonium sulphate concentration, has a surface charge that is approaching the surface charge at which wt-ocr will precipitate. Based on this observation, there must exist some critical amount of surface charge which may facilitate some specific interaction so that ocr can aggregate into ordered fibrillar

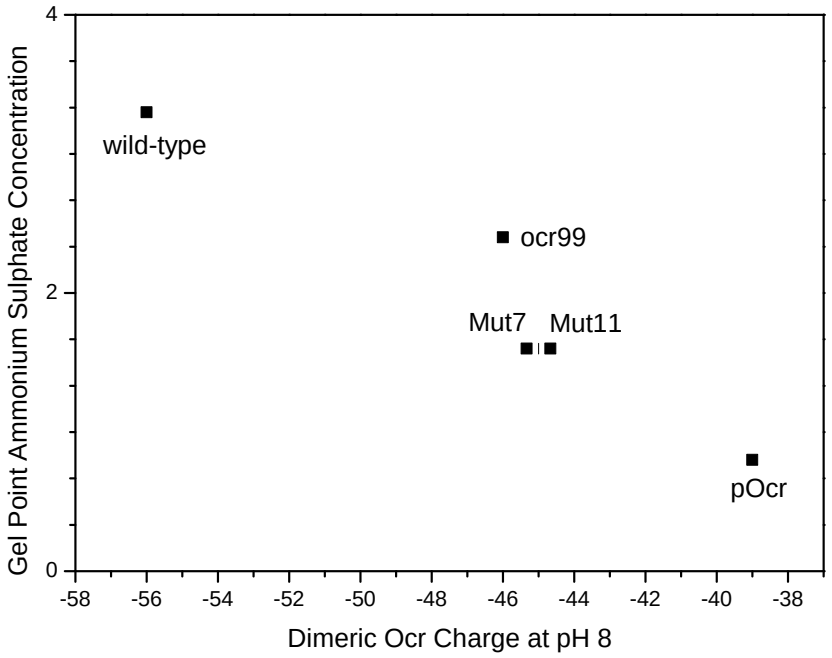

Figure 6.11: **The ammonium sulphate concentration at the gel-point scales with surface charge.** Plotted are the ammonium sulphate concentrations at which a gel is formed as a function of surface charge. One finds that the gel point ammonium sulphate concentration scales with surface charge.

structures. The formation of fine-stranded gels occurs when there is sufficient electrostatic repulsion between partially unfolded species which allows for slower and more specific association between proteins [22]. The extreme surface charge of ocr may need the excessive screening provided by such high concentrations of ammonium sulphate in order to mediate the fine balance between repulsion and making specific interactions. Of course if the charge of ocr goes below this critical surface charge, non-specific interactions dominate which results in amorphous aggregation and coagulation into particulates that can be seen by eye. However, ocr gelation must not only be related to screening alone. Indeed, at 3.2M ammonium sulphate where wt-Ocr gels, the Debye screening length is approximately 1Å; this is less than the size of a water molecule. It is then unrealistic to speak of a screening length at these salt concentrations. Ammonium

sulphate is a good salt for salting out proteins because sulphate ions occur early in the Hofmeister series. Sulphate ions are kosmotropic, that is, they allow for to better form hydrogen bonds between water molecules. Therefore, a salt not only can provide screening between other proteins but will effect the solution environment in which they exist (for instance increasing surface tension) and this ordering of water is what drives the phenomenon of 'salting out'. However, salts and their ions which come before ammonium sulphate in the Hofmeister series do not cause ocr to gel [4]. Ultimately, charge screening and Hofmeister effects are not sufficient to completely understand ocr gelation.

Many questions remain, the chief among them, is what exactly are these specific interactions which occur between ocr dimers that causes ordered fibrillar aggregates to be formed. Another important question is why is ocr fibrillar aggregation specific to ammonium sulphate and not other salts. While many different salts were tested to see whether they would induce gelation in wt-ocr, the mutants were not screened. It is conceivable that at sufficiently lowered surface charges, such as in the case of pOcr, other salts may be able to induce gel formation. Preliminary studies have shown that pOcr does form a gel at a similar concentration of NaCl to the ammonium sulphate experiments described above. This would be an interesting avenue of future work. It would also be advantageous to create additional mutants that have surface charges below this theoretical critical value. Of course it would be ideal to perform another set of experiments on ocr99 with better time resolution and statistics. It would also be useful to characterize the physical dimension of the aggregates formed from the variants. Attempts to image the aggregates using TEM were unsuccessful. Perhaps a better approach would be to use solution-state ('wet') AFM to image the aggregates. Solution-state AFM would be preferable to drying aggregates onto a substrate due to the formation of ammonium sulphate crystals. Indeed, when TEM was attempted the only visible structures on the TEM grid was ammonium sulphate. Finally, a rigorous exploration of the viscosity of the gels by microrheological methods should be performed by using a range of bead sizes to eliminate potential error introduced by the pore size of the gel. While the mechanism and organization of ocr into aggregates remains unclear, this chapter has shed light onto possible regions on the surface of ocr which may play a role in determining this unique protein aggregation phenomena.

Chapter 7

Conclusions & Outlook

7.1 Insulin Amyloid Fibrils

The kinetics of insulin amyloid fibril formation was studied using ThT fluorescence assays. The experimental protocols used in this thesis allowed for a drastic reduction in the variability in kinetic experiments (see Appendix B). Our first observation is that the kinetics match well with predictions made by a model in which fragmentation plays a dominant role in the kinetics of fibril assembly. However, we observed a transition in the kinetics at a critical protein concentration c_T. When the experimental kinetic curves are normalized, we observe a distinct plateau in the growth rates with increasing concentration for protein concentrations $c_p > c_T$. It should be emphasized that the minimization of variability allowed this observation to be made; the variability in previous experiments would have completely obscured these kinetics features.

In order to better understand the nature of this transition, the structure of the resultant fibrils were studied by CD spectroscopy. We found that the structure of the fibrils formed above and below c_T exhibit qualitatively different CD spectra. Furthermore, bead-tracking experiments demonstrated that for a transition from a fluid to a highly viscous fluid state occurs precisely at c_T. These observations lead to the postulation of a 'critical fibril concentration' (CFC). Once the fibril concentration in the system crosses this value, we hypothesize that fragmentation becomes physically arrested.

K. Eden developed computer simulations that employs the fragmentation model of [119]. This model fails to capture the transition that we observe.

However, when the simulations are modified to incorporate a CFC, we find that the model captures the features of the kinetics well. Interestingly, the CFC model predicts distinct fibril length distributions: for $c_p > c_T$ we expect a log-normal distribution populated by long fibrils; when $c_p > c_T$ fragmentation is not arrested which results in an exponential distribution of short fibrils. We measured length distributions obtained from transmission electron microscopy images. These results matched the qualitative predictions made by the CFC model.

To further probe the nature of the CFC, kinetic experiments were repeated with increasing concentrations of NaCl. We observed non-trivial and non-monotonic trends in the growth rates as a function of protein concentration (and with NaCl concentration). The features of these trends could be convincingly reproduced by the CFC model by shifting the CFC value to lower concentrations.

Short fibrils and oligomers have been implicated as being cytotoxic [184, 178, 179, 180, 181, 41]. Since fragmentation is not arrested below c_T, a very large number of small oligomeric aggregates populate the system. We speculate that the formation of stable, long amyloid fibrils may actually protect against the accumulation of these harmful species. These results point to a new picture of amyloid fibril growth in which structural transitions that occur during self-assembly have strong effects on the final population of aggregate species with small, and potentially cytotoxic, oligomers dominating for long periods of time at protein concentrations below the CFC.

In Chapter 4, we used time-resolved SANS to further interrogate the nature of the transition that we observe in the insulin system. We found that, up to a critical time, the scattering curves could be well described by a simple cylinder model, where the radius remained a constant size but the length increased from t=180 to t=245 mins. For $t \geq 270$ mins, this simple model no longer fit the data. At this critical time we observe a large increase in scattering at low-Q values indicating the presence of larger aggregates. This was further confirmed by looking at the length distribution $P(r)$. While polydispersity is an issue, we modeled the system as a simple linear combination of cylinder models. We found that the scattering signal at these longer times could be described by two populations: small, monomer-like scatterers and much longer aggregates. These longer aggregates had radii larger than the aggregates observed in the 180 mins

$\leq t \leq$=255 mins. interval. The fact that this onset of larger aggregates occur at a distinct time is intriguing and is consonant with the observations made in Chapter 3. Unfortunately, due to a major fault in the linear accelerator, we were unable to obtain complete kinetics. It would be very interesting to determine the percentage of fibril aggregates in the system, relative to initial protein concentration in these experiments. Doing so would allow a determination whether the increase in aggregate size at that critical time is coincident with the system crossing the CFC. It need not be said that further experiments need to be done in order to test this. Obtaining SANS data for insulin concentrations below c_T has also been shown to be achievable. Observing full kinetic runs for concentrations in that regime will also be highly informative and help resolve the exact mechanism which arrests fragmentation at the CFC.

7.2 Ovalbumin Fibrils

We studied the kinetics of fibril formation in reduced ovalbumin using ThT fluorescence. We found that ovalbumin displays very different kinetics from those observed in Chapter 3. The kinetics were characterized by several features. First, there was no lag time which preceded the onset of growth. Second, we observe almost immediate exponential growth. Third, this exponential growth phase does not finish in a flat plateau as we observed in insulin amyloid fibril formation in Chapter 3. Instead, there is a gradual turn-over in the kinetic curve which then proceeds nearly linearly and at a very slow rate. Furthermore, Bradford Assays indicate that the aggregation process is far from complete in these systems.

TEM was performed in order to study the morphology of the aggregates that are subsequently formed. In contrast to the long and straight fibrils we observed in Chapter 3, ovalbumin fibrils are short and worm-like. We measured a length distribution which showed that there is a large population of short fibrils ($\sim$ 200-300 nm), but was not exponential in character. Furthermore, the persistence length of the fibrils was found to be very short ($\sim$ 26 nm). One of the most striking features of the fibrils observed using TEM was the presence of closed-loop fibrils. We speculate, the fact that the persistence length is so short, facilitates two ends of a fibril to have a finite probability of meeting and joining.

We develop a simple model where fibril formation is a downhill polymerization

process. If we do not include a term which takes into account the joining of ends, the kinetics that we observe can not be well described by such a model. Namely, the simple downhill polymerization does not capture the slow growth regime *and* the time-scale of polymerization that we observe. When an end-joining term is incorporated into the model such features can capture these features. Therefore we attribute the kinetics we observe to two features. First, the lack of a lag-time is due to the system being effectively seeded. Second, the slow-growth at long times is due to end depletion due to both self-joining and fibril-to-fibril end joining.

It would be interesting in the future to attempt to isolate the fibril loops, perhaps by ultra-centrifugation. Can a system be seeded by these species? If so, what is the resulting morphology of the fibrils? Are there conditions in which these worm-like fibrils can be converted to longer and straighter fibrils? Another interesting question is why are the kinetics apparently seeded? Is it because of the solution conditions or are the fibrils nucleating at interfaces? Finally, a better characterization of the underlying structure of the fibrils needs to be attained to determine whether these fibrils are amyloid-like in nature or something else entirely.

7.3 Ocr Aggregation

Ocr forms a transparent gel at a concentration of 3.2M ammonium sulphate. The fact that ocr forms a transparent, fine-stranded gel, under such high-charge screening conditions is highly unusual. In an attempt to better understand the mechanism which causes this anomalous aggregation behavior, we studied mutant variants of ocr. These mutants had large surface charge modifications, often located in different regions of the protein. The structure of the gels formed from the mutants were interrogated using bead-tracking microrheology. These experiments revealed the concentration of ammonium sulphate needed to form a gel decreased with decreasing amounts of surface charge. This type of trend is typical of a 'salting-out' trend; however instead of precipitating ocr and its mutants form gels.

Preliminary data was taken on a chemically excised variant of ocr. This variant contained many modifications similar to those in the mutant pOcr. The

excised portion missing from ocr99 contains 10 of the 17 charge modifications of pOcr (predominately located in the unstructured C-terminus tail). It is interesting to find that the ammonium sulphate concentration at which these two variants gel are markedly different. This may be indicative that the unstructured tail (which has been removed in ocr99) does not play an important role in driving the formation of ocr fibrillar aggregates. However, a single region on the surface of the protein which strongly contributes to the assembly process can not be pin-pointed from this analysis. However, what does emerge is that charge does play a critical role in driving the formation of fibrillar aggregates. The general model that I hypothesize is that there is a critical surface charge which enables electrostatic repulsion to effectively slow the protein-protein interactions enough where specific contacts can be made. If there is not enough screening, aggregation occurs. However, if there is not enough charge on the surface, the electrostatic repulsions can not mediate the contacts between proteins, allowing for non-specific aggregation and precipitation. Much more work needs to be done to understand the nature of these interactions. The production of further mutants with greater amounts of charge removed and testing whether other salts can induce aggregation for lower surface charge mutants would be of great interest. Furthermore, studying the structure of the aggregates could shed light on the forces which hold the aggregates together.

7.4 Final Conclusion

This thesis has investigated the mechanisms of fibrillar protein aggregation in three different protein systems. Taken as a whole, these studies have provided new insights into how and why proteins self-associate into ordered fibrillar assemblies. This work has the potential to have an impact on our understanding of diseases related to fibrillar aggregates and their pathology by identifying the structural characteristics and conditions that influence the accumulation of harmful aggregates. In general, by establishing a fundamental understanding of the mechanisms of aggregation, such behavior can be controlled and possibly exploited. The applications of such knowledge are many-fold and cross many disciplines, from the improvement of protein-based pharmaceuticals to making novel and useful materials.

The approach taken in this thesis has been from a 'physics' perspective: I wished to understand, at a basic mechanistic level, the complicated and often 'messy' phenomenon of protein aggregation by employing model systems in which the parameters of the systems could be well controlled. Modeling can then be used to understand the essential forces and mechanisms of assembly by reducing these complicated systems to their minimal ingredients. Not only does modeling provide insights into the physical phenomenon uncovered in the experiments, but also offers feedback and informs new ideas for future experimental investigations. This symbiosis of experiment and modeling allows for maximum insight into the aggregation mechanisms explored in this thesis. Of course, understanding the 'real world' of *in vivo* aggregation is never so simple and the biology and biochemistry involved is highly complex. However, by taking a physics-based approach, I believe we can begin to deconvolve these highly complex systems into their essential components which can ultimately inform our understanding of the biology and chemistry of protein aggregation.

Appendix A

Temperature Inhomogeneities in Common Plate Readers: A Word of Warning

A.1 Introduction

In the course of doing a large number of kinetic experiments using a variety of plate readers, I discovered that for samples that are highly sensitive to temperature variations, some plate readers are better than others at providing homogeneous temperatures within the machines. This can have consequences for reproducibility of experiments and will increase errors and variability in kinetic data. I present a variety of common plate readers and show some are better than others at providing homogeneous temperature control. This appendix serves as a word of warning to individuals using these machines to be vigilant when performing sensitive experiments using these machines.

A.2 BMG Omega

The Omega series of plate readers made by BMG is the model on which the majority of ThT fluorescence kinetic data was obtained in this thesis. Correlations of kinetic parameters as a function of position on the plate were first noticed in this machine. Clearly, there is a correlation of the lag time from top-to-bottom in the plate where as going from left-to-right on the plate reveals no correlation

within error (Figs. A1 and A2). We noticed that the left side of the plate would sit near the door of the machine (as well as over the optic) between read cycles. The problem was fixed by programming the plate to sit far away from the door and optic. In the BMG Omega this meant positioning the plate in the far back right corner of the machine.

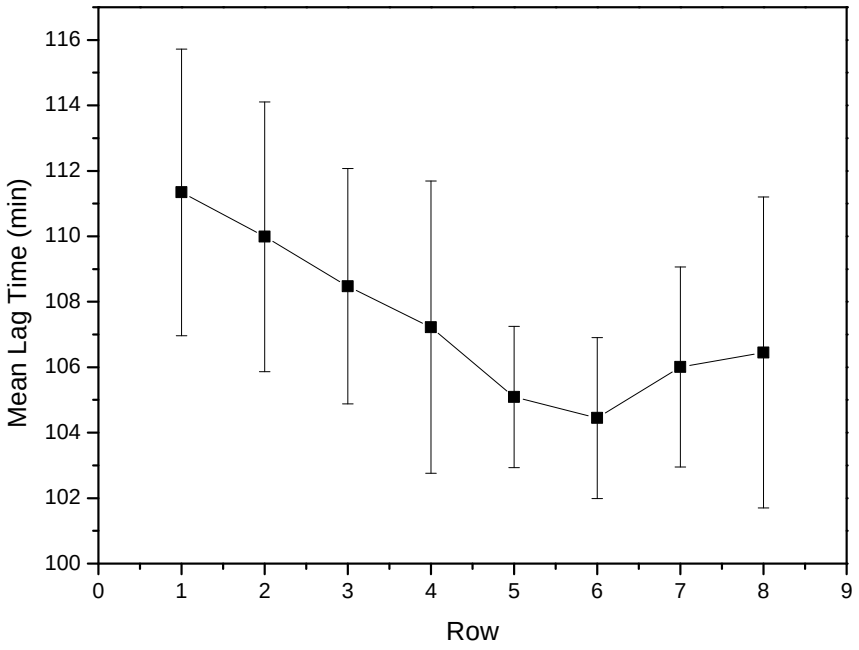

Figure A.1: **BMG Omega Mean Lag Time By Row** Displayed are the mean lag times as a function of row for a 1 mg/ml insulin sample.

A.3 Spectramax M5e

I then tested whether this sort of correlation is found in other plate readers. We tested our system on the Spectramax M5e, which is the top of the range plate reader offered by Spectramax. We again used a 1 mg/ml sample of insulin and measured the lag times. Plotted in Figs. A3 and A4 are the mean lag times as a function of row and column. The mean lag times are longer on the top of the plate relative to the bottom. Interestingly, the lag times are shorter on the left side of the plate relative to the right side. The plate sits in a different

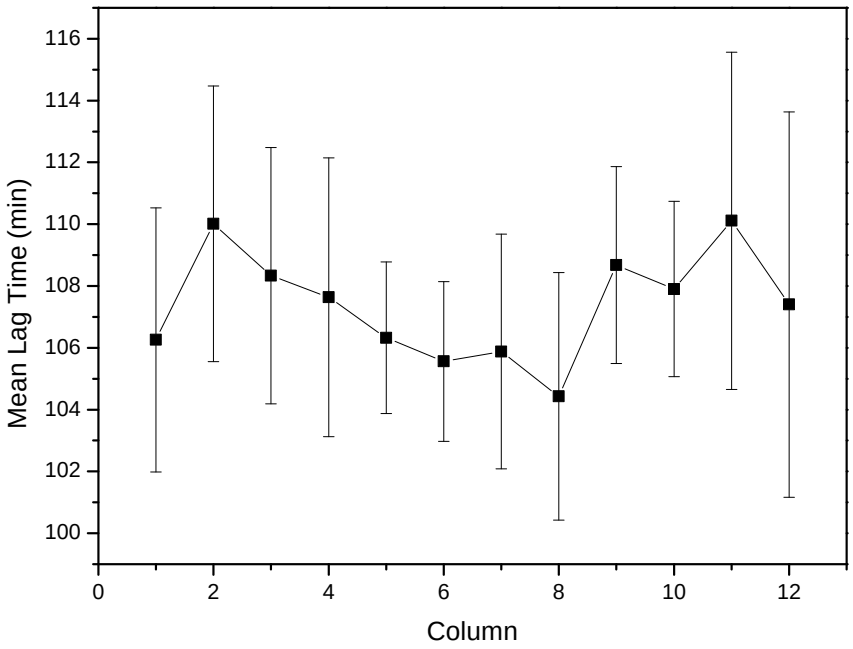

Figure A.2: **BMG Omega Mean Lag Time By Column** Displayed are the mean lag times as a function of column for a 1 mg/ml insulin sample.

orientation inside the Spectramax from the BMG. In the Spectramax, the right side of the plate sits closest to the door. We hypothesize that the doors on the BMG and Spectramax are not sealed well allowing for temperature gradients to be established within the machines.

A.4 Biotek Synergy H4

We tested two plate readers made by Biotek. The Synergy H4 is the top-of-the-line instrument offered by Biotek. When we tested the Bioteks we used a 1 mg/ml ovalbumin as our protein test system. As was outlined in Chapter 5, ovalbumin kinetics lack a lag time. In these tests we used mean maximum ThT fluorescence as the metric instead of lag time. Fig. A5 shows that the final row, that is the bottom of the plate, has a markedly decreased maximum fluorescence. This row sits closest to the door of the machine. Again, there may be sealing issues allowing for temperature gradients. The salient point gleaned from Fig. A6 is that: first,

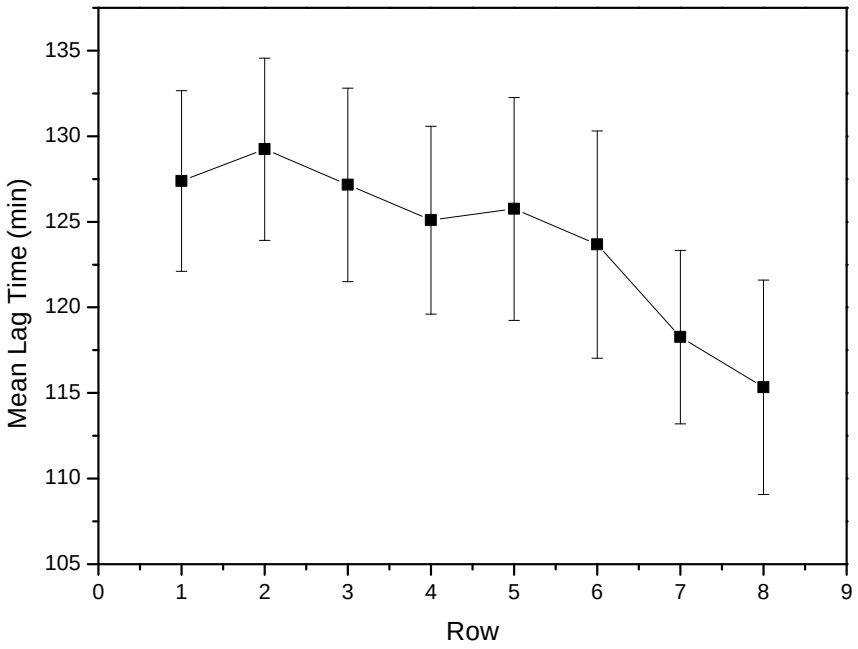

Figure A.3: **Spectramax M5e Mean Lag Time By Row** Displayed are the mean lag times as a function of row for a 1 mg/ml insulin sample.

there appears to be some correlation of the maximum fluorescence as a function of column; secondly, the last two columns have much greater variability in the mean maximum fluorescence values. These results would be of some concern if one required high precision in measuring these types of parameters.

A.5 Biotek Synergy 2

We finally tested the Biotek Synergy 2. Fig. A7 shows that there may be a slight correlation of the fluorescence as a function of the rows. In contrast, Fig. A8 demonstrates that there is no correlation as a function of column.

A.6 Conclusions

What is clear is that one must take care in performing experiments in plate readers where temperature control is of great importance. One should always

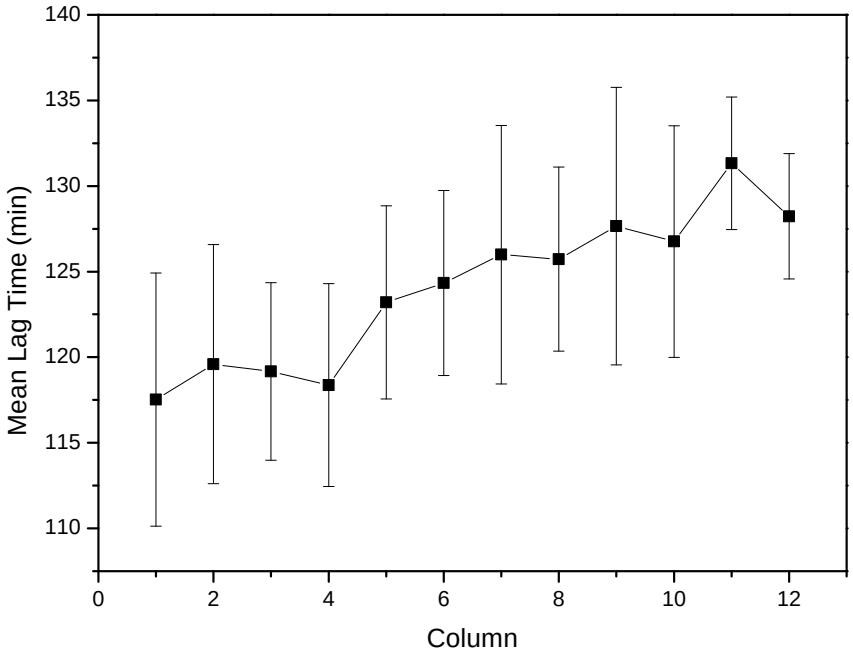

Figure A.4: **Spectramax M5e Mean Lag Time By Column** Displayed are the mean lag times as a function of column for a 1 mg/ml insulin sample.

test their results by performing correlation analyses as shown above.

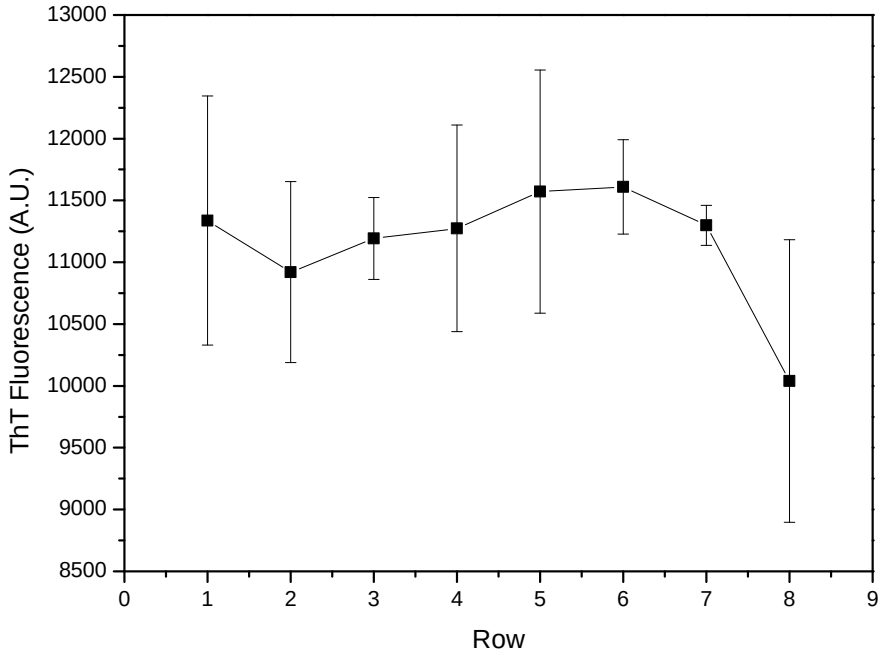

Figure A.5: **BiotekH4 Mean ThT Fluorescence By Row** Displayed is the mean maximum ThT fluorescence of 1 mg/ml ovalbumin as a function of row.

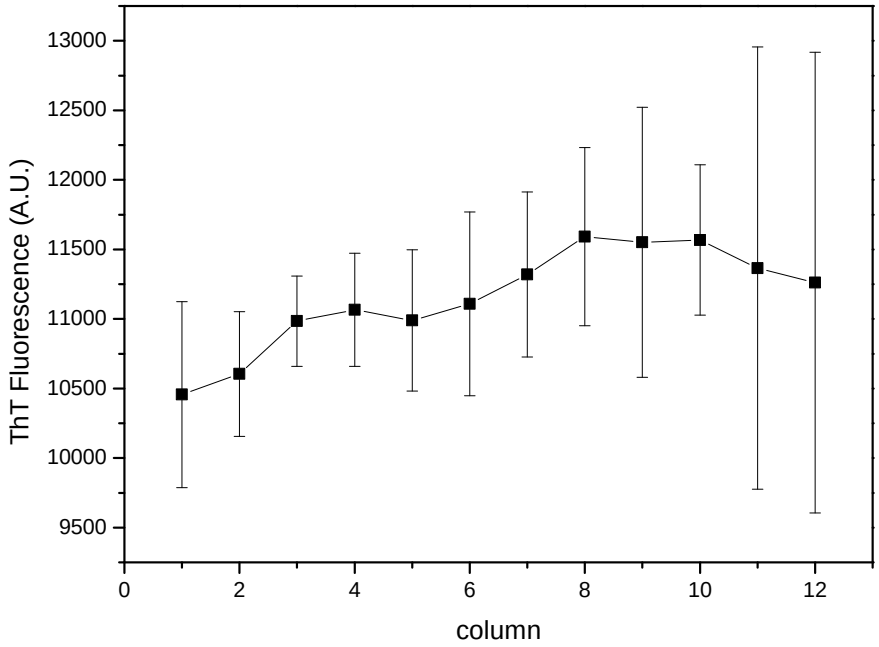

Figure A.6: **BiotekH4 Mean ThT Fluorescence By Row** Displayed is the mean maximum ThT fluorescence of 1 mg/ml ovalbumin as a function of column.

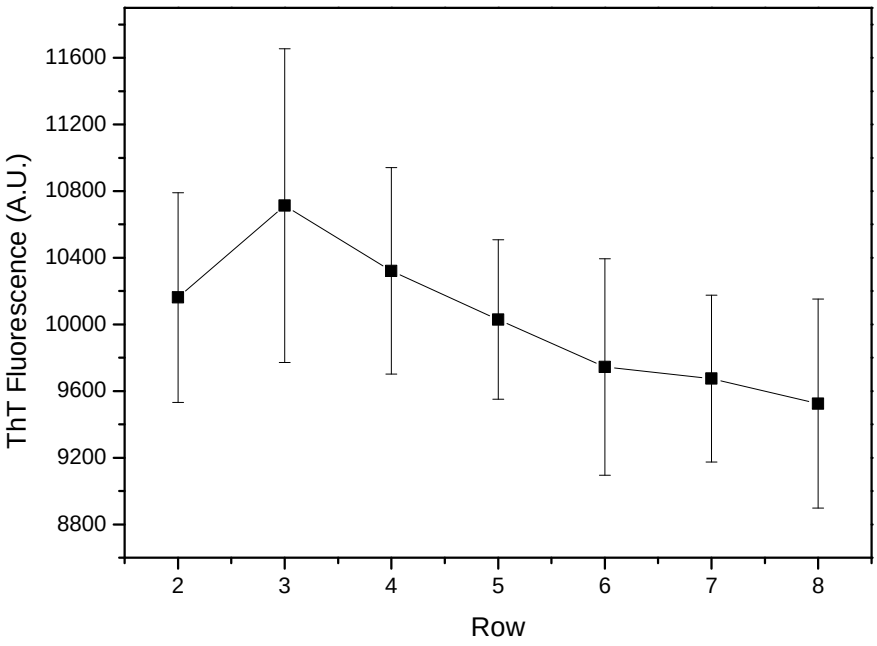

Figure A.7: **Biotek Synergy 2 Mean ThT Fluorescence By Row** Displayed is the mean maximum ThT fluorescence of 1 mg/ml ovalbumin as a function of row.

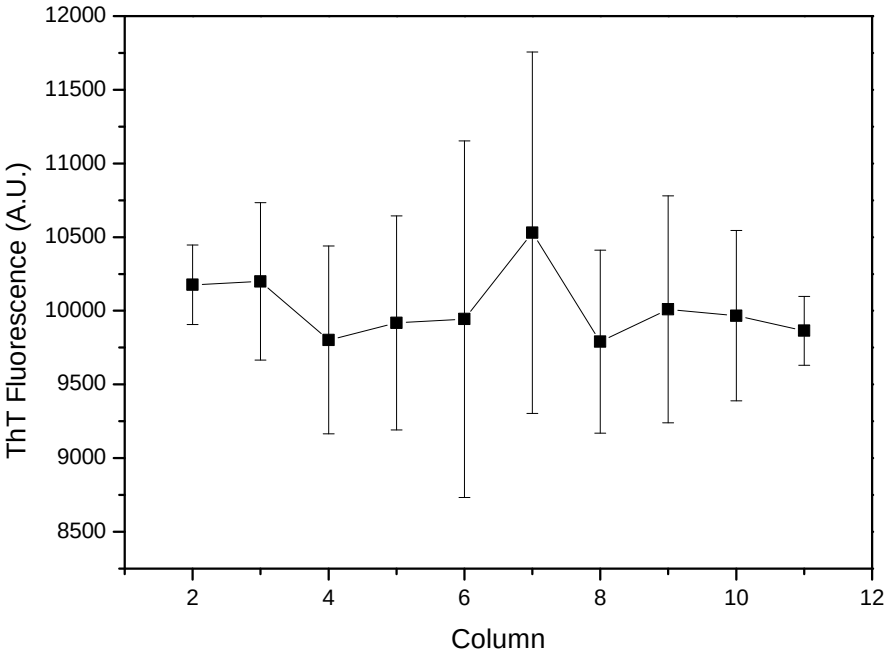

Figure A.8: **Biotek Synergy2 Mean ThT Fluorescence By Row** Displayed is the mean maximum ThT fluorescence of 1 mg/ml ovalbumin as a function of column.

Appendix B

Demonstration of Reduction of Variability in Polystyrene Plates Plates vs. NBS Plates

B.1 Polystyrene Plates: Insulin Kinetics

B.2 NBS Plates: Insulin Kinetics

B.3 Lag Time Distributions

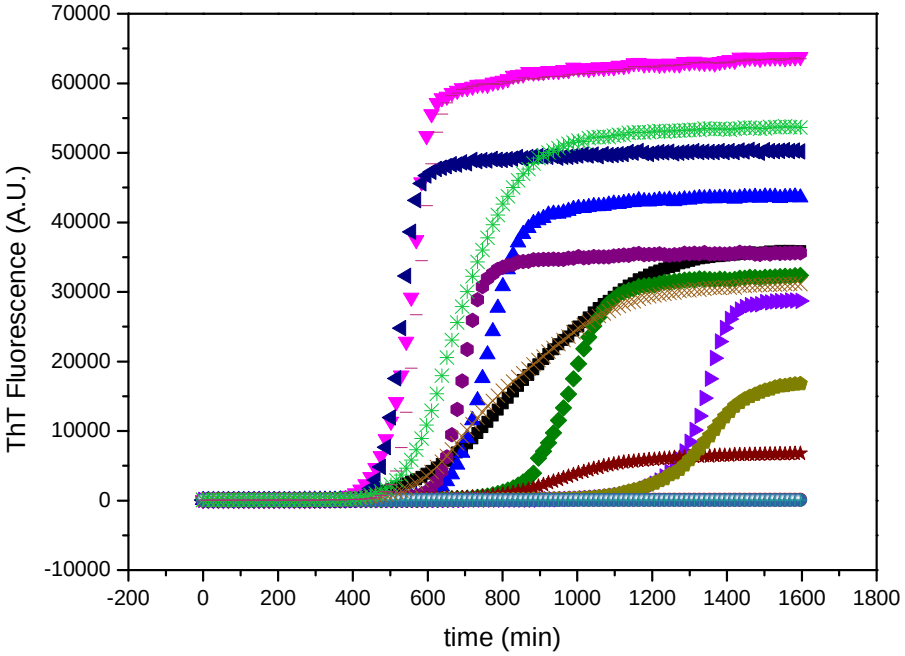

Figure B.1: **Insulin Amyloid Kinetics in Polystyrene Plate** Insulin amyloid fibril formation in polystyrene plates results in extreme variations in the kinetics from well-to-well. Shown are kinetic runs for 1 mg/ml sample.

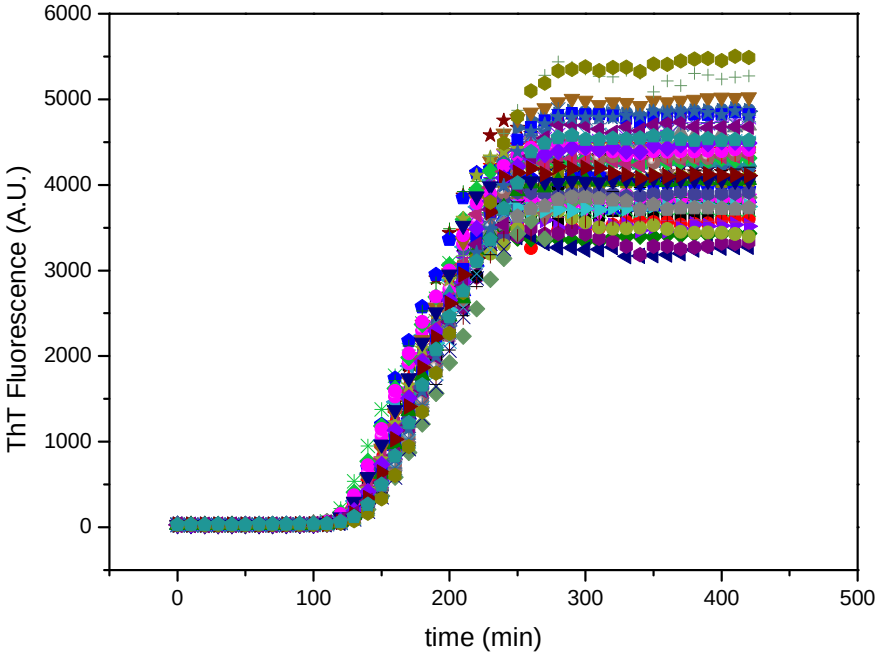

Figure B.2: **Insulin Amyloid Kinetics in NBS** Insulin amyloid fibril formation in NBS plates results in extremely well controlled kinetics with greatly minimized variation well-to-well. Shown are kinetic runs 1.5 mg/ml sample.

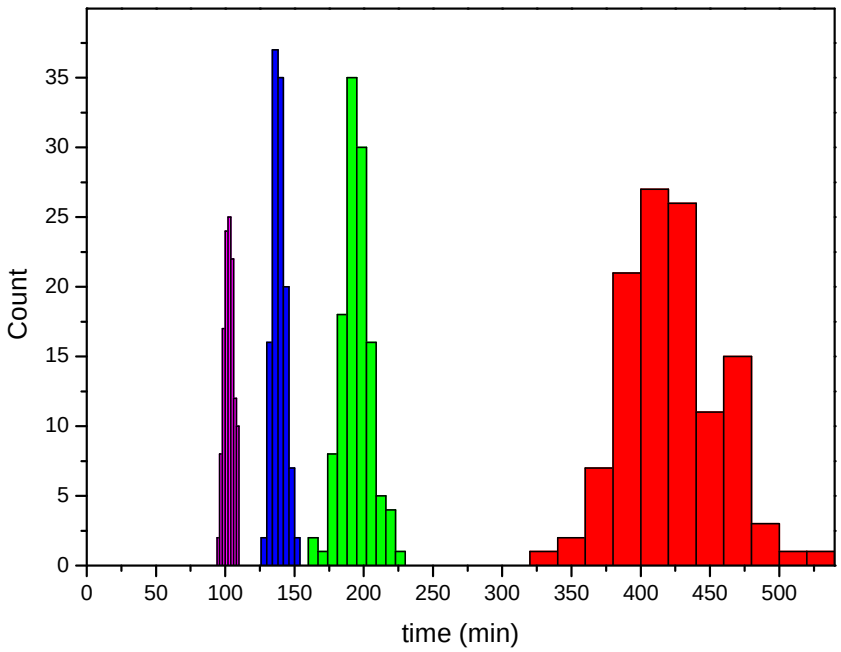

Figure B.3: **Lag Time Distributions for Insulin Amyloid Fibrils** Shown are lag time distributions for 0.1 mg/ml (red), 0.4 mg/ml (green), 1.5 mg/ml (blue), and 4 mg/ml (magenta) insulin concentrations. Note that the distributions, especially for higher concentrations are sharply peaked indicating very reproducible kinetics. This should be contrasted against what has been previously observed in insulin fibril kinetics where the variability was significantly larger [10]. Such reproducibility in our experiments has allowed for kinetic features to be observed which otherwise would be obscured by such variability.

Appendix C

Table of Kinetic Data for Agitated and Unagitated Samples

The table below is a compendium of the kinetic data obtained from both agitated and unagitated experiments.

Shaking Conditions	Conc. (mg/ml)	N	τ_{lag} (min)	s.d.	k_{fluo} (min^{-1})	s.d.	MMF (A.U.)	s.d.
Agitated	0.1	136	507.55	54.7	0.22	0.04	43.6	7.4
	0.25	136	282.05	24.18	0.6	0.15	72.36	16.3
	0.4	136	242.46	29.3	1.44	0.53	182.5	66.4
	0.75	136	196.98	10.4	12.8	4.1	1402.3	384.7
	1	136	161.48	10.5	20.9	5.1	2253.1	424.9
	1.5	136	122.42	9.85	39.54	3.8	4374.6	498.7
	2	136	97.8	12.82	59.39	10.88	5574.6	502.1
	4	136	80.45	10.8	74.09	7.97	7003.9	996.5
Unagitated	0.1	136	421.31	36.3	0.16	0.03	37.4	5.4
	0.25	136	278.06	18.06	0.51	0.15	71.1	14.5
	0.4	136	224.33	12.7	1.21	0.22	117.5	9.8
	0.75	136	181.11	6.246	2.79	0.71	271.8	54.1
	1	204	170.12	7.8	10.98	5.17	1197.3	468.1
	1.5	136	138.92	4.59	30.34	5.05	3411.9	339.7
	2	136	121.13	4.79	36.23	2.79	3587.0	5.3
	4	136	102.82	3.45	48.42	3.19	5833.5	469.8

Table C.1: **Agitated and unagitated kinetic data obtained from experiment.** N represents the number of individual wells each average data value is comprise of. MMF is the mean maximal fluorescence value and s.d. is the standard deviation.

Bibliography

[1] Jason T. Giurleo, Xianglan He, and David S. Talaga. β-lactoglobulin assembles into amyloid through sequential aggregated intermediates. *J. Mol. Biol.*, 381(5):1332–1348, 2008.

[2] Jose L. Jimenez, Ewan J. Nettleton, Mario Bouchard, Carol V. Robinson, Christopher M. Dobson, and Helen R. Saibil. The protofilament structure of insulin amyloid fibrils. *Proc Natl Acad Sci USA*, 99:9196–9201, 2002.

[3] David L. Nelson and Michael M. Cox. *Principles of Biochemistry*. W.H. Freeman and Co., 4th edition, 2007.

[4] Richard G. Higham. *A Biophysical Analysis of the Ocr Protein Gel*. PhD thesis, The University of Edinburgh, 2007.

[5] Sharon M. Kelly, Thomas J. Jess, and Nicholas C. Price. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta*, 1751:119 139, 2005.

[6] SasView, http://danse.chem.utk.edu/downloads/ModelfuncDocs.pdf. *Sas Model Function Information Manual*.

[7] D. Hall and H. Edskes. Computational modeling of the relationship between amyloid and disease. *Biophysical Reviews*, 4(3):205–222, 2012.

[8] G. Pescitelli, D. Padula, and F. Santorob. Intermolecular exciton coupling and vibronic effects in solid-state circular dichroism: a case study. *Phys. Chem. Chem. Phys.*, 15:795–802, 2013.

[9] Bruce M. Taylor, Ronald W. Sarver, Gregory Fici, Roger A. Poorman, Barry S Lutzke, Antonio Molinari, Thomas Kawabe, Karl Kappenman, Allen E. Buhl, and Dennis E. Epps. Spontaneous aggregation and cytotoxicity of the β-amyloid aβ^{1-40}: A kinetic model. *J. Protein Chem.*, 22(1):31–40, 2003.

[10] Vito Fodera, Fabio Librizzi, Minna Groenning, Marco van de Weert, and Maurizio Leone. Secondary nucleation and accessible surface in insulin amyloid fibril formation. *J. Phys. Chem. B*, 112:3853–3858, 2008.

[11] C.B. Anfinsen, E. Haber, M. Sela, and F.H. White Jr. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain". *Pr*, 47(9):13091314, 1961.

[12] C.B. Anfinsen. Principles that govern the folding of protein chains. *Science 181 (4096): 223230*, 181(4096):223230, 1973.

[13] C. Levinthal. Are there pathways for protein folding? *Journal de Chimie Physique et de Physico-Chimie Biologique*, 65:4445, 1968.

[14] C. Levinthal. How to fold graciously. In *Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois*, 1969.

[15] W.A. Eaton, V. Munoz, S.J. Hagen, G.S. Jas, L.J. Lapidus, E.R. Henry, and J. Hofrichter. Fast kinetics and mechanisms in protein folding. *Annual review of biophysics and biomolecular structure*, 29(1):327–359, 2000.

[16] D.B. Wetlaufer. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci USA*, 70:697–701, 1973.

[17] R.L. Baldwin. The nature of protein folding pathways: the classical view versus the new view. *J. Biomolec. NMR*, 5:103–109, 1995.

[18] K.A. Dill and H.S. Chan. From levinthal to pathways to funnels. *Nature Struct. Biol.*, 4:10–19, 1997.

[19] Fabrizio Chiti and Christopher M. Dobson. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.*, 75:333–366, 2006.

[20] R. Wetzel. Mutations and off-pathway aggregation of proteins. *Trends in Biotechnology*, 12(5):193–198, 1994.

[21] T.C. Squier. Oxidative stress and protein aggregation during biological aging. *Exp. Gerontol.*, 36(9):1539–50, 2001.

[22] Mark R. H. Krebs, Glyn L. Devlin, and A. M. Donald. Protein particulates: Another generic form of protein aggregation? *Biophys J.*, 92(4):1336–1342, 2007.

[23] J. Hofrichter. Kinetics of sickle hemoglobin polymerization. iii: Nucleation rates determined from stochastic fluctuations in polymerization progress curves. *Journal of Molecular Biology*, 189:553–571, 1986.

[24] LGB Bremer, Ton van Vliet, and Pieter Walstra. Theoretical and experimental study of the fractal nature of the structure of casein gels. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phase*, 85(10):3359–3372, 1989.

[25] David S Horne. Formation and structure of acidified milk gels. *International Dairy Journal*, 9(36):261–268, 1999.

[26] T. Lefevre and M. Subirade. Molecular differences in the formation and structure of fine-stranded and particulate beta-lactoglobulin gels. *Biopolymers*, 54(7):578–586, 2000.

[27] Anne-Marie Hermansson, Ole Harbitz, and Maud Langton. Formation of two types of gels from bovine myosin. *Journal of the Science of Food and Agriculture*, 37:1:69–84, 1986.

[28] S. Ikeda and V.J. Morris. Fine-stranded and particulate aggregates of heat-denatured whey proteins visualized by atomic force microscopy. *Biomacromolecules*, 3(2):382–389, 2002.

[29] E. Allen Foegeding. Food biophysics of protein gels: A challenge of nano and macroscopic proportions. *FOBI*, 1:41–50, 2006.

[30] Marleen Verheul and Sebastianus P. F. M. Roefs. Structure of particulate whey protein gels: Effect of nacl concentration, ph, heating temperature, and protein composition. *J. Agric. Food Chem.*, 46 (12):49094916, 1998.

[31] F.U. Hartl and M. Hayer-Hartl. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science*, 295:1852–1858, 2002.

[32] Craig Hammond and Ari Helenius. Quality control in the secretory pathway: Retention of a misfolded viral membrane glycoprotein involves cycling between the er, intermediate compartment, and golgi apparatus. *The Journal of Cell Biology*, 126(1):41–52, 1994.

[33] Vittorio Bellotti and Monica Stoppini. Protein misfolding diseases. *The Open Biology Journal*, 2:228–234, 2009.

[34] Mark C. Manning, Kamlesh Patel, and Ronald T. Borchardt. Stability of protein pharmaceuticals. *Pharmaceutical Research*, 6(11):903–918, 1989.

[35] Carl J. Burke, Bryan L. Steadman, David B. Volkin, Pei Kuo Tasi, Mark W. Bruner, and C. Russel Middaugh. The adsorption of proteins to pharmaceutical container surfaces. *International Journal of Pharmaceutics*, 86(1):89–93, 1992.

[36] Mary E. M. Cromwell, Eric Hilario, and Fred Jacobson. Protein aggregation and bioprocessing. *The AAPS Journal*, 8(3):572–579, 2006.

[37] Eva Y. Chi, Sampathkumar Krishnan, Theodore W. Randolph, and John F. Carpenter. Physical stability of proteins in aqueous solution: Mechanism and driving forces, in nonnative protein aggregation. *Pharmaceutical Research*, 20:1325–1336, 2003.

[38] Wei Wang. Instability, stabilization, and formulation of liquid protein pharmaceuticals. *International Journal of Pharmaceutics*, 1(2):1–30, 2005.

[39] Liza Nielsen, Ritu Khurana, Alisa Coats, Sven Frokjaer, Jens Brange, Sandip Vyas, Vladimir N. Uversky, and Anthony L. Fink. Effect of environmental factors on the kinetics of insulin fibril formation: Elucidation of the molecular mechanism. *Biochemistry*, 40:6036–6046, 2001.

[40] Monica Bucciantini, Elisa Giannoni, Fabrizio Chiti, Fabiana Baroni, Lucia Formigli, Jesus Zurdo, Niccolo Taddei, Giampietro Ramponi, Christopher M. Dobson, and Massimo Stefani. Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature*, 416:507–511, 2002.

[41] Charles G. Glabe and Rakez Kayed. Common structure and toxic function of amyloid oligomers implies a common mechanism of pathogenesis. *Neurology*, 66(1):S74–S78, 2006.

[42] Arjan Quist, Ivo Doudevski, Hai Lin, Rushana Azimova, Douglas Ng, Blas Frangione, Bruce Kagan, Jorge Ghiso, and Ratnesh Lal. Amyloid ion channels: A common structural link for protein-misfolding disease. *Proc Natl Acad Sci USA*, 102(30):104271043, 2005.

[43] H.A. Lashuel & D. Hartley, B.M. Petre, T. Walz, and P.T. Lansbury. Neurodegenerative disease: amyloid pores from pathogenic mutations. *Nature*, 418:291–292, 2002.

[44] B Caughey and PT Lansbury. Protofibrils, pores, fibrils, and neurodegeneration: Separating the responsible protein aggregates from the innocent bystanders. *Annual Review of Neuroscience*, 26:26798, 2003.

[45] E. D. Eanes and G. G. Glenner. X-ray diffraction studies on amyloid filaments. *J. Histochem. Cytochem*, 16:673–677, 1968.

[46] Mark R. H. Krebs, Eric H. C. Bromley, and Athene M. Donald. The binding of thioflavin-t to amyloid fibrils: Localisation and implications. *Journal of Structural Biology*, 149:30–37, 2005.

[47] Minna Groenning, Lars Olsen, Marco van de Weert, James M. Flink, Sven Frokjaer, and Flemming S. Jorgensen. Study of the binding of thioflavin-t to beta-sheet-rich and non-beta-sheet cavities. *Journal of Structural Biology*, 158:358–369, 2007.

[48] Minna Groenning, Mathias Norman, James M. Flink, Marco van de Weert, Jens T. Bukrinsky, Gerd Schluckebier, and Sven Frokjaer. Binding mode of thioflavin-t in insulin amyloid fibrils. *Journal of Structural Biology*, 159:483–497, 2007.

[49] James D. Harper and Peter T. Lansbury. Model of amyloid seeding in alzheimer's disease and scrapie: Mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *Annu. Rev. Biochem.*, 66:385–407, 1997.

[50] Aaron K. Chamberlain, Cait E. MacPhee, Jess Zurdo, Ludmilla A. Morozova-Roche, H. Allen O. Hill, Christopher M. Dobson, and Jason J. Davis. Ultrastructural organization of amyloid fibrils by atomic force microscopy. *Biophysical Journal*, 79:3282–3293, 2000.

[51] Jose L. Jimenez, J. Inaki Guijarro, Elena Orlova, Jesus Zurdo, Christopher M.Dobson, Margaret Sunde, and Helen R.Saibil. Cryo-electron microscopy structure of an sh3 amyloid fibril and model of the molecular packing. *The EMBO Journal*, 18:815821, 1999.

[52] Walraj S. Gosal, Isobel J. Morten, Eric W. Hewitt, D. Alastair Smith, Neil H. Thomson, and Sheena E. Radford. Competing pathways determine fibril morphology in the self-assembly of β2-microglobulin into amyloid. *J. Mol. Biol.*, 351:850–864, 2005.

[53] A. Aggeli, I. A. Nyrkova, M. Bell, R. Harding, L. Carrick, T. C. B. McLeish, A. N. Semenov, and N. Boden. Hierarchical self-assembly of chiral rod-like molecules as a model for peptide β-sheet tapes, ribbons, fibrils, and fibers. *Proc Natl Acad Sci USA*, 98(21):11857–11862, 2001.

[54] Mario Bouchard, Jesus Zurdo, Ewan J. Nettleton, Christopher M. Dobson, and Carol V. Robinson. Formation of insulin amyloid fibrils followed by ftir simultaneously with cd and electron microscopy. *Protein Science*, 9:10:1960–1967, 2000.

[55] Aneta T. Petkova, Yoshitaka Ishii, John J. Balbach, Oleg N. Antzutkin, Richard D. Leapman, Frank Delaglio, and Robert Tycko. A structural model for alzheimers beta-amyloid fibrils based on experimental constraints from solid state nmr. *Proc Natl Acad Sci USA*, 99:16742–16747, 2002.

[56] Danny M. Hatters, Cait E. MacPhee, Lynne J. Lawrence, William H. Sawyer, and Geoffrey J. Howlett. Human apolipoprotein c-ii forms twisted amyloid ribbons and closed loops. *Biochemistry*, 39 (28):82768283, 2000.

[57] Danny M. Hatters, Christopher A. MacRaild, Rob Daniels, Walraj S. Gosal, Neil H. Thomson, Jonathan A. Jones, Jason J. Davis, Cait E. MacPhee, Christopher M. Dobson, and Geoffrey J. Howlett. The circularization of amyloid fibrils formed by apolipoprotein c-ii. *Biophys. J.*, 85(6):3979–3990, 2003.

[58] Jesper Sndergaard Pedersen, Dantcho Dikovb, James L. Flinkc, Hans Aage Hjulerb, Gunna Christiansend, and Daniel Erik Otzen. The changing face of glucagon fibrillation: Structural polymorphism and conformational imprinting. *J. Mol. Biol.*, 355(3):501523, 2006.

[59] Fabrizio Chiti, Niccol Taddei, Fabiana Baroni, Cristina Capanni, Massimo Stefani, Giampietro Ramponi, and Christopher M. Dobson. Kinetic partitioning of protein folding and aggregation. *Nature Structural Biology*, 9:137 – 143, 2002.

[60] C Wurth, NK Guimard, and MH Hecht. Mutations that reduce aggregation of the alzheimer's a [beta] 42 peptide: An unbiased search for the sequence determinants of a [beta] amyloidogenesis. *Journal of Molecular Biology*, 319(5):12791290, 2002.

[61] Daniel E. Otzen, Ole Kristensen, and Mikael Oliveberg. Designed protein tetramer zipped together with a hydrophobic alzheimer homology: A structural clue to amyloid assembly. *Proc*, 97(18):99079912, 2000.

[62] Russell Schwartz, Sorin Istrail, and Jonathan King. Frequencies of amino acid strings in globular protein sequences indicate suppression of blocks of consecutive hydrophobic residues. *Protein Sci.*, 10(5):10231031, 2001.

[63] Fabrizio Chiti, Martino Calamai, Niccol Taddei, Massimo Stefani, Giampietro Ramponi, and Christopher M. Dobson. Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. *Proc Natl Acad Sci USA*, 99:1641926, 2002.

[64] Yvonne Kallberg, Magnus Gustafsso, Bengt Persson, Johan Thyberg, and Jan Johansson. Prediction of amyloid fibril-forming proteins. *Journal of Biological Chemistry*, 276:12945–12950, 2001.

[65] Thomas R. Jahn and Sheena E. Radford. The yin and yang of protein folding. *FEBS Journal*, 272:59625970., 2005.

[66] Sally L. Gras, Anna K. Tickler, Adam M. Squires, Glyn L. Devlin, Michael A. Horton, Christopher M. Dobson, and Cait E. MacPhee. Functionalised amyloid fibrils for roles in cell adhesion. *Biomaterials*, 29:1553–1562, 2008.

[67] Morten S. Dueholm, Steen V. Petersen, Mads Snderkr, Poul Larsen, Gunna Christiansen, Kim L. Hein, Jan J. Enghild, Jeppe L. Nielsen, Kre L. Nielsen, Per H. Nielsen, and Daniel E. Otzen1. Functional amyloid in pseudomonas. *Molecular Microbiology*, 77:10091020, 2010.

[68] Douglas M. Fowler, Atanas V. Koulov, William E. Balch, and Jeffery W. Kelly. Functional amyloid from bacteria to humans. *TRENDS in Biochemical Sciences*, 32:217–224, 2007.

[69] Douglas M. Fowler, Atanas V. Koulov, Christelle Alory-Jost, Michael S. Marks, William E. Balch, and Jeffery W. Kelly. Functional amyloid formation within mammalian tissue. *PLoS Biology*, 4:100–107, 2006.

[70] Tuomas P. J. Knowles and Markus J. Buehler. Nanomechanics of functional and pathological amyloid materials. *Nature Nanotechnology*, 6:469–479, 2011.

[71] Tuomas P. Knowles, Anthony W. Fitzpatrick, Sarah Meehan, Helen R. Mott, Michele Vendruscolo, Christopher M. Dobson, and Mark E. Welland. Role of intermolecular forces in defining material properties of protein nanofibrils. *Science*, 318:1900–1903, 2007.

[72] Natsuki Kobayashi, Sungwoong Han, Chikashi Nakamura, and Koji Sode. Nanostructure fabrication based on engineeredα-synuclein. *Nanobiotechnol.*, 4:50–55, 2008.

[73] Timothy S. Burkoth, Tammie L. S. Benzinger, Volker Urban, David G. Lynn, Stephen C. Meredith, and P. Thiyagarajan. Self-assembly of aβ(10-35)-peg block copolymer fibrils. *J. Am. Chem. Soc.*, 121:7429–7430, 1999.

[74] Joel H. Collier and Phillip B. Messersmith. Self-assembling polymer peptide conjugates: nanostructural tailoring. *Advanced Materials*, 16:907–910, 2004.

[75] Shuguang Zhang. More than just bare scaffolds: towards multi-component and decorated fibrous biomaterials. *Nature Biotechnology*, 21:1171–1178, 2003.

[76] Thomas Scheibel, Jesse Bloom, and Susan L. Lindquist. The elongation of yeast prion fibers involves separable steps of association and conversion. *Proc. Natl. Acad. Sci. USA.*, 101(8):22872292, 2004.

[77] Meital Reches and Ehud Gazit. Casting metal nanowires within discrete self-assembled peptide nanotubes. *Science*, 300:625–627, 2003.

[78] Ohad Carny, Deborah E. Shalev, and Ehud Gazit. Fabrication of coaxial metal nanocables using a self-assembled peptide nanotube scaffold. *Nano Lett.*, 6:1594–1597, 2006.

[79] Shuguang Zhang, Todd C. Holmes, C. Michael DiPersio, Richard 0. Hynes, Xing Su, and Alexander Rich. Self-complementary oligopeptide matrices support mammalian cell attachment. *Biomaterials*, 16:1385–1393, 1995.

[80] Xiumei Wang, Akihiro Horii, and Shuguang Zhang. Designer functionalized self-assembling peptide nanofiber scaffolds for growth, migration, and tubulogenesis of human umbilical vein endothelial cells. *Soft Matter*, 4:23882395, 2008.

[81] Lisa Haines-Butterick, Karthikan Rajagopal, Monica Branco, Daphne Salick, Ronak Rughani, Matthew Pilarz, Matthew S. Lamm, Darrin J. Pochan, and Joel P. Schneider. Controlling hydrogelation kinetics by peptide design for three-dimensional encapsulation and injectable delivery of cells. *Proc Natl Acad Sci USA*, 104:77917796, 2007.

[82] Todd C. Holmes, Sonsoles de Lacalle, Xing Su, Guosong Liu, Alexander Rich, and Shuguang Zhang. Extensive neurite outgrowth and active synapse formation on self-assembling peptide scaffolds. *PNAS*, 97:67286733, 2000.

[83] Sarah M. Pilkington, Sarah J. Roberts, Susie J. Meade, and Juliet A. Gerrard. Amyloid fibrils as a nanoscaffold for enzyme immobilization. *Biotechnol. Prog.*, 26:93–100, 2010.

[84] Christoph Meier and Mark E. Welland. Wet-spinning of amyloid protein nanofibers into multifunctional high-performance biofibers. *Biomacromolecules*, 12:3453–3459, 2011.

[85] Ghibom Bhak, Soonkoo Lee, Jae W. Park, Sunghyun Cho, and Seung R. Paik. Amyloid hydrogel derived from curly protein fibrils of /*alpha*-synuclein. *Biomaterials*, 31:5986–5995, 2010.

[86] Kevin J. Channon, Glyn L. Devlin, and Cait E. MacPhee. Efficient energy transfer within self-assembling peptide fibers: A route to light-harvesting nanomaterials. *J. Am. Chem. Soc.*, 131:12520–12521, 2009.

[87] Kevin J. Channon, Glyn L. Devlin, Steven W. Magennis, Chris E. Finlayson, Anna K. Tickler, Carlos Silva, and Cait E. MacPhee. Modification of fluorophore photophysics through peptide-driven self-assembly. *J. Am Chem. Soc.*, 130:5487–5491, 2008.

[88] Yeon Sun Choi, Jehoon Kim, Ghibom Bhak, Daekyun Lee, and Seung R. Paik. Photoelectric protein nanofibrils of α-synuclein with embedded iron and phthalocyanine tetrasulfonate. *Angew. Chem.*, 123:6194 6198, 2011.

[89] P. Hammarstrom, X. Jiang, A.R. Hurshman, E. T. Powers, and J. W. Kelly. Sequence-dependent denaturation energetics: a major determinant in amyloid disease diversity. *Proc Natl Acad Sci USA*, 99:16427–16432, 2002.

[90] Atta Ahmad, Ian S. Millet, Sebastian Doniach, Vladimir N. Uversky, and Anthony L. Fink. Partially folded intermediates in insulin fibrillation. *Biochemistry*, 42:11404–11416, 2003.

[91] Jeanette Otte, Richard Ipsen, Rogert Bauer, Morten J. Bjerrum, and Rianne Waninge. Formation of amyloid-like fibrils upon limited proteolysis of bovine alpha-lactalbumin. *International Dairy Journal*, 15:219–229, 2005.

[92] M. Fandrich. Absolute correlation between lag time and growth rate in the spontaneous formation of several amyloid-like aggregates and fibrils. *J. Mol. Biol.*, 365:1266–1270, 2007.

[93] Christian Rischel Fabio Librizzi. The kinetic behavior of insulin fibrillation is determined by heterogeneous nucleation pathways. *Protein Science*, 14:3129–3134, 2005.

[94] Aleksey Lomakin, Doo Soo Chung, George B. Benedek, Daniel A. Kirschner, and David B. Teplow. On the nucleation and growth of amyloid beta-protein fibrils: Detection of nuclei and quantitation of rate constants. *Proc Natl Acad Sci USA*, 93:1125–1129, 1996.

[95] Aleksey Lomakin, David B. Teplow, Daniel A. Kirschner, and George B. Benedek. Kinetic theory of fibrillogenesis of amyloid beta-protein. *Proc Natl Acad Sci USA*, 94:7942–7947, 1997.

[96] Yoko Kusumoto, Aleksey Lomakin, David B. Teplow, and George B. Benedek. Temperature dependence of amyloid beta-protein fibrillization. *Proc Natl Acad Sci USA*, 95:12277–12282, 1998.

[97] Mauro Manno, Emanuela Fabiola Craparo, Vincenzo Martorana, Donatella Bulone, and Pier Luigi San Biagio. Kinetics of insulin aggregation: Disentanglement of amyloid fibrillation from large-size cluster formation. *Biophysical Journal*, 90:4585–4591, 2006.

[98] M. I. Smith, J. S. Sharp, and C. J. Roberts. Insulin fibril nucleation: The role of prefibrillar aggregates. *Biophysical Journal*, 95:3400–3406, 2008.

[99] David P. Smith, Sheena E. Radford, and Alison E. Ashcroft. Elongated oligomers in β_2-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry. *Proc Natl Acad Sci USA*, 107:6794–6798, 2010.

[100] Ewan J. Nettleton, Paula Tito, Margaret Sunde, Mario Bouchard, Christopher M. Dobson, and Carol V. Robinson. Characterization of the oligomeric states of insulin in self-assembly and amyloid fibril formation by mass spectrometry. *Biophys. J.*, 79:1053–1065, 2000.

[101] Atta Ahmad, Vladimir N. Uversky, Dongpyo Hong, and Anthony L. Fink. Early events in the fibrillation of monomeric insulin. *Journal of Biological Chemistry*, 280:42669–42675, 2005.

[102] Dominic M. Walsh, Dean M. Hartley, Yoko Kusumoto, Youcef Fezoui, Margaret M. Condron, Aleksey Lomakin, George B. Benedek, Dennis J. Selkoe, and David B. Teplow. Amyloid beta-protein fibrillogenesis: Structure and biological activity of protofibrillar intermediates. *The Journal of Biological Chemistry*, 274:25945–25952, 1999.

[103] Simone Hess, Susan L. Lindquist, and Thomas Scheibel. Alternative assembly pathways of the amyloidogenic yeast prion determinant sup35-nm. *EMBO Reports*, 8:1196–1201, 2007.

[104] Alessandro Podesta, Guido Tiana, Paolo Milani, and Mauro Manno. Early events in insulin fibrillization studied by time-lapse atomic force microscopy. *Biophysical Journal*, 90:589–597, 2006.

[105] Ralf Jansen, Wojciech Dzwolak, and Roland Winter. Amyloidogenic self-assembly of insulin aggregates probed by high resolution atomic force microscopy. *Biophysical Journal*, 88:1344–1353, 2005.

[106] Neil M. Kad, Sarah L. Myers, David P. Smith, D. Alastair Smith, Sheena E. Radford, and Neil H. Thomson. Hierarchical assembly of β_2-microglobulin amyloid in vitro revealed by atomic force microscopy. *J. Mol. Biol.*, 330:785–797, 2003.

[107] Ritu Khurana, Cristian Ionescu-Zanetti, Maighdlin Pope, Jie Li, Liza Nielsen, Marina Ramirez-Alvarado, Lynn Regan, Anthony L. Fink, and Sue A. Carter. A general model for amyloid fibril assembly based on morphological studies using atomic force microscopy. *Biophysical Journal*, 85:1135–1144, 2003.

[108] Hisashi Yagi, Tadato Ban, Kenichi Morigaki, Hironobu Naiki, and Yuji Goto. Visualization and classification of amyloid-beta supramolecular assemblies. *Biochemistry*, 46:15009–15017, 2007.

[109] Shae B. Padrick and Andrew D. Miranker. Islet amyloid: Phase partioning and secondary nucleation are central to the mechanism of fibrillogenesis. *Biochemistr*, 41:4694–4703, 2002‘.

[110] Bente Vestergaard, Minna Groenning, Manfred Roessle, Jette S. Kastrup, Marco van de Weert, James M. Flink, Sven Frokjaer, Michael Gajhede, and Dmitri I. Svergun. A helical stuctural nucleus is the primary elongating unit of insulin amyloid fibrils. *PLoS Biology*, 5:1089–1097, 2007.

[111] Winnie Yong, Aleksey, Lomakin, Marina D. Kirkitadze, David B. Teplow, Sow hsin Chen, and George B. Benedek. Structure determination of micelle-like intermediates in amyloid β-protein fibril assembly by using small angle neutron scattering. *Proc Natl Acad Sci USA*, 99:150–154, 2002.

[112] Susan Krueger, Derek Ho, and Amos Tsai. *Misbehaving Proteins*, chapter Small-Angle Neutron Scattering as a Probe for Protein Aggregation at Many Length Scales. Springer New York, 2006.

[113] Robert Tycko. Biomolecular solid state nmr: Advances in structural methodology and applications to peptide and protein fibrils. *Annu. Rev. Phys. Chem*, 52:575–606, 2001.

[114] Christopher P. Jaroniec, Cait E. MacPhee, Vikram S. Bajaj, Michael T. McMahon, Christopher M. Dobson, and Robert G. Griffin. High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning nmr spectroscopy. *Proc Natl Acad Sci USA*, 101:711–716, 2004.

[115] Frank Ferrone. Analysis of protein aggregation. *Methods in Enzymology*, 309:256–273, 1999.

[116] F. Oosawa and M. Kasai. A theory of linear and helical aggregations of macromolecules. *Journal of Molecular Biology*, 4:10–21, 1962.

[117] Henrik Flyvbjerg, Elmar Jobs, and Stanislas Leibler. Kinetics of self-assembling microtubules: An 'inverse problem' in biochemistry. *Proc Natl Acad Sci USA*, 93:5975–5979, 1996.

[118] F.A. Ferrone, J. Hofrichter, and W. A. Eaton. Kinetics of sickle hemoglobin polymerization. ii: A double nucleation mechanism. *Journal of Molecular Biology*, 183:611–631, 1985.

[119] Tuomas P. J. Knowles, Christopher A Waudby, Glyn L. Devlin, Samuel I. A. Cohen, Adriano Aguzzi, Michele Vendruscolo, Eugene M. Terentjev, Mark E. Welland, and Christopher M. Dobson. An analytical solution to the kinetics of breakable filament assembly. *Science*, 326:1533–1537, 2009.

[120] Samuel I. A. Cohen, Michele Vendruscolo, Mark E. Welland, Christopher M. Dobson, Eugene M. Terentjev, and Tuomas P. J. Knowles. Nucleated polymerization with secondary pathways. i. time evolution of the principal moments. *J. Chem. Phys.*, 135:65–105, 2011.

[121] S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles. Nucleated polymerization with secondary pathways. ii. determination of self-consistent

solutions to growth processes described by non-linear master equations. *J. Chem. Phys.*, 135:65–105, 2011.

[122] S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles. Nucleated polymerization with secondary pathways. iii. equilibrium behavior and oligomer populations. *J. Chem. Phys.*, 135, 2011.

[123] Peter Hortchansky, Volker Schroeckh, Tony Christopeit, Giorgia Zandomeneghi, and Marcus Fandrich. The aggregation kinetics of alzheimer's β-amyloid peptide is controlled by stochastic nucleation. *Protein Science*, 14:1753–1759, 2005.

[124] Zhiqi Cao and Frank Ferrone. A 50th order reaction predicted and observed for sickle hemoglobin nucleation. *J. Mol. Biol.*, 256:219–222, 1996.

[125] A. Szabo. Fluctuations in the polymerization of sickle hemoglobin: A simple analytic model. *J. Mol. Biol.*, 199(3):539–542, 1988.

[126] Erik Hellstrand, Barry Boland, Dominic M. Walsh, and Sara Linse. Amyloid β-protein aggregation produces highly reproducible kinetic data and occurs by a two-phase process. *ACS Chem. Neurosci.*, 1:13–18, 2010.

[127] Tuomas P. J. Knowles, Duncan A. White Adam R. Abate, Jeremy J. Agresti, Samuel I. A. Cohen, Ralph A. Sperling, Erwin J. De Genst, Christopher M. Dobson, and David A. Weitz. Observation of spatial propagation of amyloid assembly from single nuclei. *Proc. Natl. Acad. Sci. USA.*, 108 (36):14746–14751, 2011.

[128] Allen H. Pekar and Bruce H. Frank. Conformation of proinsulin. a comparison of insulin and proinsulin self-association at neutral ph. *Biochemistry*, 11(22):4013–4016, 1972.

[129] Jean L Whittingham, David J Scott, Karen Chance, Ashley Wilson, John Finch, Jens Brange, and G Guy Dodson. Insulin at ph 2: Structural analysis of the conditions promoting insulin fibre formation. *J. Mol. Biol.*, 318(2):479–490, 2002.

[130] Lauge Schaffer. A model for insulin binding to the insulin receptor. *European Journal of Biochemistry*, 221(3):11271132, 1994.

[131] Jean L. Whittingham, David J. Scott, Karen Chance, Ashley Wilson, John Finch, Jens Brange, and G. Guy Dodson. Insulin at ph 2: Structural analysis of the conditions promoting insulin fibre formation. *J. Mol. Biol.*, 318:479–490, 2002.

[132] Liza Nielsen, Sven Frokjaer, Jens Brange, Vladimir N. Uversky, and Anthony L. Fink. Probing the mechanism of insulin fibril formation with insulin mutants. *Biochemistry*, 40:8397–8409, 2001.

[133] D.F. Waugh. A fibrous modification of insulin. i. the heat precipitate of insulin. *J. Am. Chem. Soc.*, 68:247 250, 1946.

[134] D.F. Waugh, D.F. Wilhelmson, S.L. Commerford, and M.L. Sackler. Studies of the nucleation and growth reactions of selected types of insulin fibrils. *J. Am. Chem. Soc.*, 75:2592 2600, 1953.

[135] Victoria Sluzky, Janet A. Tamada, Alexander M. Klibanov, and Robert Langer. Kinetics of insulin aggregation in aqueous solutions upon agitation in the presence of hydrophobic surfaces. *Proc Natl Acad Sci USA*, 88:9377–9381, 1991.

[136] Vito Fodera, Sebastiano Cataldo, Fabio Librizzi, Bruno Pignataro, Paola Spiccia, and Maurizio Leone. Self-organization pathways and spatial heterogeneity in insulin amyloid fibril formation. *J. Phys. Chem. B*, 113:10830–10837, 2009.

[137] Vito Fodera, Marco van de Weert, and Bente Vestergaard. Large-scale polymorphism and auto-catalytic effect in insulin fibrillogenesis. *Soft Matter*, 6:4413–4419, 2010.

[138] Dong-Pyo Hong, Atta Ahmad, and Anthony L. Fink. Fibrillation of human insulin a and b chains. *Biochemistry*, 45:9342–9353, 2006.

[139] Manno Mauro, Emanuela Fabiola Craparo, Alessandro Podesta, Donatella Bulone, Rita Carrotta, Vincenzo Martorana, Guido Tiana, and Pier Luigi San Biagio. Kinetics of difference processes in human insulin amyloid formation. *J. Mol. Biol.*, 366:358–274, 2007.

[140] M. I. Smith, J. S. Sharp, and C. J. Roberts. Nucleation and growth of insulin fibrils in bulk solution and at hydrophobic polystyrene surfaces. *Biophysical Journal*, 93:2143–2151, 2007.

[141] F. E. Dische, C. Wernstedt, G. T. Westermark, P. Westermark, M. B. Pepys, J. A. Rennie, S. G. Gilbey, and P. J. Watkins. Insulin as an amyloid-fibril protein at sites of repeated insulin injections in a diabetic patient. *Diabetologia*, 31(3):158–161, 1988.

[142] James A. Huntington and Penelope E. Stein. Structure and properties of ovalbumin. *Journal of Chromatography B: Biomedical Sciences and Applications*, 756, Issues 12:189–198, 2001.

[143] Akio Kato and Toshio Takagi. Formation of intermolecular α-sheet structure during heat denaturation of ovalbumin. *J. Agric. Food Chem.*, 36:1156–1159, 1988.

[144] Naoki Tanaka, Yumi Morimoto, Yurika Noguchi, Tomoko Tada, Tomonori Waku, Shigeru Kunugi, Takashi Morii, Yin-Fai Lee, Takashi Konno, and Nobuyuki Takahashi. The mechanism of fibril formation of a non-inhibitory serpin ovalbumin revealed by the identification of amyloidogenic core regions. *The Journal of Biological Chemistry*, 286(7):58845894, 2011.

[145] Reza Khodarahmi, Mehdi Beyrami, and Hosnieh Soor. Appraisal of caseins inhibitory effects on aggregation accompanying carbonic anhydrase refolding and

heat-induced ovalbumin fibrillogenesis. *Archives of Biochemistry and Biophysics*, 477:6776, 2008.

[146] David T. F. Dryden. Dna mimicry by proteins and the control of enzymatic activity on dna. *Trends in Biotechnology*, 24:378–382, 2006.

[147] C. Atanasiu, O. Byron, H. McMiken, S. S. Strurrock, and D. T. F. Dryden. Characterization of the structure of ocr, the gene 0.3 protein of bacteriophage t7. *Nucleic Acids Research*, 29 (14):3059–3068, 2001.

[148] M.D. Walkinshaw, P. Taylor, S. S. Sturrock, C. Atanasiu, T. Berge, R. M. Henderson, J. M. Edwardson, and D. T. F. Dryden. Structure of ocr from bacteriophage t7, a protein that mimics b-form dna. *Molecular Cell*, 9:187–194, 2002.

[149] A. S. Stephanou, G. A. Roberts, L. P. Cooper, D. J. Clarke, A. R. Thompson, C. L. MacKay, M. Nutley, A. Cooper, and D. T. F. Dryden. Dissection of the dna mimicry of the bacteriophage t7 ocr protein using chemical modification. *J. Mol. Biol.*, 391(3):565–576, 2009.

[150] H. Naiki, K. Higuchi, M. Hosokawa, and T. Takeda. Fluorometric determination of amyloid fibrils in vitro using the fluorescent dye, thioflavin t. *Anal. Biochem.*, 177:244–249, 1989.

[151] H. LeVine. Stopped-flow kinetics reveal multiple phases of thioflavin t binding to alzheimer β(1-40) amyloid fibrils. *Arch. Biochem. Biophys.*, 342:306–316, 1997.

[152] T. G. Mason and D. A. Weitz. Optical measurements of frequency-dependent linear viscoelastic moduli of complex fluids. *Phys.Rev.Lett.*, 74(7):1250–1253, 1995.

[153] F. Amblard, A.C. Maggs, B. Yurke, A. N. Pargellis, and S. Leibler. Subdiffusion and anomalous local viscoelasticity in actin networks. *Phys. Rev. Lett.*, 77(21):4470–4473, 1996.

[154] John C. Crocker and David G. Grier. Methods of digital video microscopy for colloidal studies. *J. Colloid Interface Sci.*, 179(298), 1996.

[155] Robert W. Woody. Circular dichroism. *Methods in Enzymology*, 246:34–71, 1995.

[156] A. Lobley, L. Whitmore, and B.A. Wallace. Dichroweb: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics*, 18:211 212, 2002.

[157] Peter B. Moore. Small-angle scattering. information content and error analysis. *J. Appl. Cryst.*, 13:168–175, 1980.

[158] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):23402361, 1977.

[159] Margaret Sunde, Louise C. Serpell, Mark Bartlam, Paul E. Fraser, Mark B. Pepys, and Colin C. F. Blake. Common core structure of amyloid fibrils by synchrotron x-ray diffraction. *J. Mol. Biol.*, 273:729–739, 1997.

[160] Christiane Ritter, Marie-Lise Maddelein, Ansgar B. Siemer, Thorsten Lhrs, Matthias Ernst, Beat H. Meier, Sven J. Saupe, and Roland Riek. Correlation of structural elements and infectivity of the het-s prion. *Nature*, 435:84448, 2005.

[161] Hangyu Zhang, Hanlin Luo, and Xiaojun Zhao. Mechanistic study of self-assembling peptide rada16-i in formation of nanofibers and hydrogels. *Journal of Nanotechnology in Engineering and Medicine*, 1:011007(1–6), 2010.

[162] Meital Reches and Ehud Gazit. Controlled patterning of aligned self-assembled peptide nanotubes. *Nature Nanotechnology*, 1:195–200, 2006.

[163] R.J.B. Morris and C.E. MacPhee. Amyloid protein biomaterials. In *Encylopedia of Biophysics*. Springer-Verlag, 2012.

[164] Evan T. Powers and David L. Powers. The kinetics of nucleated polymerizations at high concentrations: Amyloid fibril formation near and above the supercritical concentration. *Biophys. J.*, 94(1):122132, 2006.

[165] Chuang-Chung Lee, Arpan Nayak, Ananthakrishnan Sethuraman, Georges Belfort, and Gregory J. McRae. A three-stage kinetic model of amyloid fibrillation. *Biophys. J.*, 92(10):34483458, 2007.

[166] Monica M. Pallitto and Regina M. Murphy. A mathematical model of the kinetics of β-amyloid fibril growth from the denatured state. *Biophysical J.*, 81(3):18051822, 2001.

[167] Aimee M. Morris, Murielle A. Watzky, Jeffrey N. Agar, and Richard G. Finke. Fitting neurological protein aggregation kinetic data via a 2-step, minimal/ockham's razor model: The finke-watzky mechanism of nucleation followed by autocatalytic surface growth. *Biochemistry*, 47(8):24132427, 2008.

[168] Aimee M. Morris, Murielle A. Watzky, and Richard G. Finke. Protein aggregation kinetics, mechanism, and curve-fitting: A review of the literature. *BBA: Proteins and Proteomics*, 1794(3):375397, 2009.

[169] Evan T. Powers and David L. Powers. Mechanisms of protein fibril formation: Nucleated polymerization with competing off-pathway aggregation. *Biophys. J.*, 94(2):379391, 2008.

[170] Wei-Feng Xue, Steve W. Homans, and Sheena E. Radford. Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proc Natl Acad Sci USA*, 105:8926–8931, 2008.

[171] Salman S. Rogers, Paul Venema, Leonard M. C. Sagis, Erik van der Linden, and Athene M. Donald. Measuring the length distribution of a fibril system: A flow birefringence technique applied to amyloid fibrils. *Macromolecules*, 38(7):29482958, 2005.

[172] Andrew J. Baldwin, Spencer J. Anthony-Cahill, Tuomas P.J. Knowles, Guy Lippens, John Christodoulou, Paul D. Barker, and Christopher M. Dobson. Measurement of amyloid fibril length distributions by inclusion of rotational motion in solution nmr diffusion measurements. *Angewandte Chemie*, 47(18):33853387, 2008.

[173] Mirco Sorci, Robert A. Grassucci, Ingrid Hahn, Joachim Frank, and Georges Belfort. Time-dependent insulin oligomer reaction pathway prior to fibril formation: Cooling and seeding. *Proteins: Structure, Function, and Bioinformatics*, 77(1):62–763, 2009.

[174] J. Iaki Guijarro, Margaret Sunde, Jonathan A. Jones, Iain D. Campbell, and Christopher M. Dobson. Amyloid fibril formation by an sh3 domain. *Proc Natl Acad Sci USA*, 95(8):4224–4228, 1998.

[175] Raul Perez-Jimenez, Raquel Godoy-Ruiz, Beatriz Ibarra-Molero, and Jose M. Sanchez-Ruiz. The efficiency of different salts to screen charge interactions in proteins: A hofmeister effect? *Biophysical Journal*, 86(4):24142429., 2004.

[176] Alexander K. Buell, Jamie R. Blundell, Christopher M. Dobson, Mark E. Welland, Eugene M. Terentjev, and Tuomas P. J. Knowles. Frequency factors in a landscape model of filamentous protein aggregation. *PRL*, 104:228101(1–4), 2010.

[177] Philipp Neudecker, Paul Robustelli, Andrea Cavalli, Patrick Walsh, Patrik Lundstrm, Arash Zarrine-Afsar, Simon Sharpe, and Lewis E. Kay. Structure of an intermediate state in protein folding and aggregation. *Science*, 336(6079):362–366, 2012.

[178] Rakez Kayed, Elizabeth Head, Jennifer L. Thompson, Theresa M. McIntire, Saskia C. Milton, Carl W. Cotman, and Charles G. Glabe. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science*, 200(5618):486–489, 2003.

[179] Sylvain Lesné, Ming Teng Koh, Linda Kotilinek, Rakez Kayed, Charles G. Glabe, Austin Yang, Michela Gallagher, and Karen H. Ashe. A specific amyloid-β protein assembly in the brain impairs memory. *Nature*, 440:352–357, 2005.

[180] Ganesh M Shankar, Shaomin Li, Tapan H Mehta, Amaya Garcia-Munoz, Nina E Shepardson, Imelda Smith, Francesca M Brett, Michael A Farrell, Michael J Rowan, Cynthia A Lemere, Ciaran M Regan, Dominic M Walsh, Bernardo L Sabatini, and Dennis J Selkoe. Amyloid-β protein dimers isolated directly from alzheimer's brains impair synaptic plasticity and memory. *Nature Medicine*, 14:837–842, 2008.

[181] Ivo Cristiano Martins, Inna Kuperstein, Hannah Wilkinson, Elke Maes, Mieke Vanbrabant, Wim Jonckheere, Patrick Van Gelder, Dieter Hartmann, Rudi D'Hooge, Bart De Strooper, Joost Schymkowitz, and Frederic Rousseau. Lipids revert inert a-beta amyloid fibrils to neurotoxic protofibrils that affect learning in mice. *EMBO Journal*, 27:224–233, 2008.

[182] A Campbell. Beta-amyloid: friend or foe. *Med Hypotheses*, 56(3):388–391, 2001.

[183] HG Lee, G Casadesus, X Zhu, JA Joseph, G Perry, and MA Smith. Perspectives on the amyloid-beta cascade hypothesis. *J. Alzheimers Dis.*, 6(2):137–145, 2004.

[184] Wei-Feng Xue, Andrew L. Hellewell, Walraj S. Gosal1, Steve W. Homans, Eric W. Hewitt, and Sheena E. Radford. Fibril fragmentation enhances amyloid cytotoxicity. *J. Bio. Chem.*, 284:34272–34282, 2009.

[185] RD Terry. The pathogenesis of alzheimer disease: an alternative to the amyloid hypothesis. *J Neuropathol Exp Neurol*, 55:10231025, 1996.

[186] Mónica Mendes Sousa, Isabel Cardoso, Rui Fernandes, António Guimarães, and Maria João Saraiva. Deposition of transthyretin in early stages of familial amyloidotic polyneuropathy: Evidence for toxicity of nonfibrillar aggregates. *Am J. Pathol.*, 159(6):19932000, 2001.

[187] CA McLean, RA Cherny, FW Fraser, SJ Fuller, MJ Smith, K Beyreuther, AI Bush, and CL Masters. Soluble pool of abeta amyloid as a determinant of severity of neurodegeneration in alzheimer's disease. *Ann. Neurol.*, 46(6):860–866, 1999.

[188] Howard Jay Aizenstein, Robert D. Nebes, Judith A. Saxton, Julie C. Price, Chester A. Mathis, Nicholas D. Tsopelas, Scott K. Ziolko, Jeffrey A. James, Beth E. Snitz, Patricia R. Houck, Wenzhu Bi, Ann D. Cohen, Brian J. Lopresti, Steven T. DeKosky, Edythe M. Halligan, and William E. Klunk. Frequent amyloid deposition without significant cognitive impairment among the elderly. *Arch. Neurol.*, 65(11):1509–1517, 2008.

[189] AY Hsia, E Masliah, L McConlogue, and *et al.* Plaque-independent disruption of neural circuits in alzheimers disease mouse models. *Proc Natl Acad Sci USA*, 96:32283233, 1999.

[190] MA Westerman, D Cooper-Blacketer, A Mariash, and *et al.* The relationship between aβ and memory in the tg2576 mouse model of alzheimers disease. *J Neurosci*, 22:18581867, 2002.

[191] D. Iacono, W. R. Markesbery, M. Gross, O. Pletnikova, G. Rudow, P. Zandi, and J. C. Troncoso. The nun study: Clinically silent ad, neuronal hypertrophy, and linguistic skills in early life. *Neurology*, 73(9):665–673, 2009.

[192] Alexander K. Buell, Jamie R. Blundell, Christopher M. Dobson, Mark E. Welland, Eugene M. Terentjev, and Tuomas P. J. Knowles. Frequency factors in a landscape model of filamentous protein aggregation. *Phys. Rev. Lett.*, 104(22):228101(1–4), 2010.

[193] M. R. Krebs, E. H. C. Bromley, S. S. Rogers, and A. M. Donald. The mechanism of amyloid spherulite formation by bovine insulin. *Biophysical Journal*, 88:2013–2021, 2005.

[194] M.R. Krebs, C.E. MacPhee, A.F. Miller, I.E. Dunlop, C.M. Dobson, and A.M. Donald. The formation of spherulites by amyloid fibrils of bovine insulin. *Proc Natl Acad Sci USA*, 101(40):14420–4, 2004.

[195] Caryn L. Heldt, Mirco Sorci, David Posada, Amir Hirsa, and Georges Belfort. Detection and reduction of microaggregates in insulin preparations. *Biotechnology and Bioengineering*, 108:237–241, 2011.

[196] Caryn L. Heldt, Dmitry Kurouski, Mirco Sorci, Elizabeth Grafeld, Igor K. Lednev, and Georges Belfort. Isolating toxic insulin amyloid reactive species that lack b-sheets and have wide ph stability. *Biophysical Journal*, 100:27922800, 2011.

[197] Arpan Nayak, Mirco Sorci, Susan Krueger, and Georges Belfort. A universal pathway for amyloid nucleus and precursor formation for insulin. *Proteins: Structure, Function, and Bioinformatics*, 74(3):556565, 2009.

[198] Christopher B. Stanley, Tatiana Perevozchikova, and Valerie Berthelier. Structural formation of huntingtin exon 1 aggregates probed by small-angle neutron scattering. *Biophysical Journal*, 100:25042512, 2011.

[199] Guy Dodson and Don Steiner. The role of assembly in insulin's biosynthesis. *Current Opinion in Structural Biology*, 8:189–194, 1998.

[200] Vincent Zoetea, Markus Meuwly, and Martin Karplus. A comparison of the dynamic behavior of monomeric and dimeric insulin shows structural rearrangements in the active monomer. *J. Mol. Biol.*, 342(3):913929, 2004.

[201] Harriet Cole, Massimiliano Porrini, Callum Davidson, Tom Smith, Jason Kalapothakis, Ryan Morris, Stefan Weidt, C. Logan Mackay, Cait E. MacPhee, and Perdita E. Barran. Early stages of insulin fibrillogenesis detailed with mass spectrometry and molecular modelling. *JACS*, (In Press), 2013.

[202] Leonard F. Pease, Mirco Sorci, Suvajyoti Guha, De-Hao Tsai, Michael R. Zachariah, Michael J. Tarlov, and Georges Belfort. Probing the nucleus model for oligomer formation during insulin amyloid fibrillogenesis. *Biophysical Journal*, 99(12):3979–3985, 2010.

[203] Julia S. Higgins and Henri C. Benoit. *Polymers and Neutron Scattering*. Oxford University Press, 1994.

[204] Claudio Rivetti, Martin Guthold, and Carlos Bustamante. Scanning force microscopy of dna deposited onto mica: Equilibration versus kinetic trapping studied by statistical polymer chain analysis. *J. Mol. Biol.*, 264:919932, 1996.

[205] Jeffrey F. Smith, Tuomas P. J. Knowles, Christopher M. Dobson, Cait E. MacPhee, and Mark E. Welland. Characterization of the nanoscale properties of individual amyloid fibrils. *Proc Natl Acad Sci USA*, 103:15806–15811, 2006.

[206] Sarah Meehan, Tuomas P. J. Knowles, Andrew J. Baldwin, Jeffrey F. Smith, Adam M. Squires, Phillip Clements, Teresa M. Treweek, Heath Ecroyd, Gian Gaetano Tartaglia, Michele Vendruscolo, Cait E. MacPhee, Christopher M. Dobson, and John A. Carver. Characterisation of amyloid fibril formation by small heat-shock chaperone proteins human αa-, αb- and r120g αb-crystallins. *J. Mol. Biol.*, 372:470484, 2007.

[207] K.A. Conway, S.J. Lee, J.-C. Rochet, T.T. Ding, R.E. Williamson, and P.T. Lansbury Jr. Acceleration of oligomerization, not fibrillization, is a shared property of both a-synuclein mutations linked to early-onset parkinsons disease: implications for pathogenesis and therapy. *Proc. Natl. Acad. Sci. USA.*, 97:571576, 2000.

[208] Nikolaj K. Holm, Stine K. Jespersen, Lise V. Thomassen, Tine Y. Wolff, Pankaj Sehgal, Line A. Thomsen, Gunna Christiansen, Christian Beyschau Andersen, Anders D. Knudsen, and Daniel E. Otzen. Aggregation and fibrillation of bovine serum albumin. *Biochimica et Biophysica Acta*, 1774:11281138, 2007.

[209] A. Stephanou. *A Biophysical Study of the DNA Charge Mimicry Displayed by the T7 Ocr Protein.* PhD thesis, University of Edinburgh, 2010.

Publications

Ryan J. Morris, Kym Eden, Reuben Yarwood, Line Jourdain, Rosalind J. Allen & Cait E. MacPhee. Mechanistic and environmental control of the prevalence and lifetime of amyloid oligomers. In *Nature Commun.*, 2013.

Ryan J. Morris, Cait E. MacPhee. Amyloid Protein Biomaterials. In *Encylopedia of Biophysics*, Springer-Verlag, 2012.